AFIT/GE/ENG/95D-23

Maximum Likelihood Estimation of Wave Front Slopes using a Hartmann-type Sensor

THESIS

Scott Anthony Sallberg
Captain, USAF

AFIT/GE/ENG/95D-23

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

AFIT/GE/ENG/95D-23

Maximum Likelihood Estimation of Wave Front Slopes using a Hartmann-type Sensor

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Electrical Engineering

Scott Anthony Sallberg, BSEE

Captain, USAF

December 1995

Preface

The Problem

> *Twinkle, twinkle little star, How I wonder what you are,*
> *Up above the world so high, Like a diamond in the sky...* [38]

The Observation

> *If the Theory of making Telescopes could at length be fully brought into Practice, yet there would be certain Bounds beyond which Telescopes could not perform. For the Air through which we look upon the Stars, is in perpetual Tremor; as may be seen by the tremulous Motion of Shadows cast from high Towers, and by the twinkling of the fix'd Stars. But these Stars do not twinkle when viewed through Telescopes which have large apertures. For the Rays of Light which pass through divers parts of the aperture, tremble each of them apart, and by means of their various and sometimes contrary Tremors, fall at one and the same time upon different points in the bottom of the Eye, and their trembling Motions are too quick to be perceived severally. And all these illuminated Points constitute one broad lucid Point, composed of those many trembling Points confusedly and insensibly mixed with one another by very short and swift Tremors, and thereby cause the Star to appear broader than it is, and without any trembling of the whole. Long Telescopes may cause Objects to appear brighter and larger than short ones can do, but they cannot be so formed as to take away that confusion of the Rays which arises from the Tremors of the Atmosphere. The only Remedy is a most serene and quiet Air, such as may perhaps be found on the tops of the highest Mountains above the grosser Clouds* [26].

A Solution

Sir Isaac Newton's keen observation provides an insightful introduction to the fundamental limitations of astronomical imaging imposed by atmospheric turbulence. In recent times, automated systems have been employed to sharpen the image of a celestial object before it is recorded [23] and computer software has been developed to mitigate the blurring effects of atmospheric turbulence [21]. The technique developed in this thesis seeks to characterize the cumulative effects of the atmosphere between the celestial body and adaptive optical telescope [36]. Specifically, this technique estimates wave front slopes over each subaperture of a Hartmann-type wave front sensor.

Scott Anthony Sallberg

Acknowledgements

While the responsibility for completing this thesis rested squarely on my shoulders, I didn't do it alone, I received assistance, guidance, and encouragement from many people. Without the love and support of my wife, Captain Kristen Gray Sallberg, this thesis would have been more work than play, more business than pleasure, more tedium than fun–thanks K. Dr Byron M. Welsh's guidance, enthusiasm for the project, and expert insight were instrumental in making this thesis more than a lengthy collection of words and equations; Major Michael C. Roggemann and Captain Joseph J. Sacchini added thoughtful words of encouragement and their time; and First Lieutenant Donald A. McCandless and Captain Eric E. Silbaugh gave their friendship and hours of enlightened discussion. My two-year old daughter, Elisabeth Eileen Sallberg, instructed me in the subtle art of differentiating wants and needs with her timely plea, "*play with me Daddy.*"

Scott Anthony Sallberg

Table of Contents

List of Figures

Abstract

Current methods for estimating the wave front slope at the pupil of a telescope equiped with a Hartmann-type wave front sensor (H-WFS) are based on a simple centroid calculation of the intensity distributions (spots) recorded in each subaperture of the H-WFS. The centroid method does not include any knowledge concerning correlation properties of the slopes over the subapertures or the amount of light collected by the telescope and diverted to the H-WFS for wave front reconstruction purposes. This thesis devises a maximum likelihood (ML) estimation of the spot centroids by incorporating statistical knowledge of the spot shifts. The light level in each subaperture and the relative spot size is also employed by the shift estimator. The shift estimator is found to be unbiased and is upper bounded by the mean squared error performance exhibited by the classical centroid technique.

Maximum Likelihood Estimation of Wave Front Slopes using a Hartmann-type Sensor

I. Background

1.1 Introduction

Many years have passed since Isaac Newton recorded his observations of the "Tremors of the Atmosphere" [26]. These "Tremors" are responsible for the scintillation effects commonly called twinkling. We now know that the diurnal heating and cooling of the Earth's surface is the root cause of atmospheric turbulence [32]. This turbulence reduces the effective resolution of a telescope and consequently blurs imagery of celestial objects [8].

Although atmospheric turbulence is unavoidable, the distortion effects (scintillation and blur) can be reduced by proper site selection [32]. Placing an observatory on top of a mountain, where the air is "serene and quiet," can lessen the deleterious effects of the atmosphere. For example, the Air Force Maui Optical Station (AMOS) is located atop the 10,000-ft Mount Haleakala on the island of Maui, Hawaii [2]. Unfortunately, this "Remedy" proposed by Newton, falls short of achieving the theoretical resolution of large diameter telescopes. Although larger diameter telescopes allow one to view dimmer objects, they do not significantly improve the resolution. In recent years, adaptive optical (AO) imaging systems have been successfully employed to compensate for the distortions induced by the atmosphere [46].

AO imaging systems compensate for atmospheric effects before the image is formed by "realigning" the light rays. Without adaptive optics, the theoretical resolution achieved by modern telescopes is on par with those of the amateur astronomer–only a couple of arc seconds [32]. In 1953, Horace Babcock suggested that the atmospheric blurring of astronomical imagery could be reduced by mechanical means–an AO imaging system [17]. In

1956, Robert Leighton, a self-proclaimed amateur astronomer, recorded the clearest images of Jupiter, Saturn, and Mars published at that time [23]. Leighton attached a "simple" first-order AO system to the 60-inch reflecting telescope at the Mount Wilson Observatory.

AO imaging systems technology is an important area of current research aimed at improving image quality [1]. The performance of an AO imaging system is fundamentally limited by many factors. The most significant factors are the finite amount of light diverted to the wave front sensor (WFS) and the finite spatial sampling of the incident wave front by the WFS [36]. Cost also limits the performance of AO imaging systems. As the cost of these sophisticated systems increases, many organizations and amateur astronomers may turn to hybrid approaches. A hybrid approach may be used to supplement deficiencies in an AO imaging system by combining mechanical pre-compensation from an AO system and image post-processing [27, 34].

Image post-processing techniques range from the simple inverse filter [13, 18] to sophisticated blind deconvolution methods [19, 22, 41]. A relatively recent hybrid technique known as deconvolution from wave front sensing (DWFS) was proposed by Fried [9] in 1987 and extended by Primot *et al.* [31] in 1990. DWFS explicitly uses WFS data from an imaging system to improve image resolution.

The Air Force mission of imaging exo-atmospheric objects provides the central motivation for developing techniques to overcome the deletrious effects of the atmosphere on ground-based imaging [6].

The purpose of the following sections is to illustrate the importance of estimating the wave front slope using a Hartmann-type WFS (H-WFS) for wavefront reconstruction purposes. The first pertinent background topic discussed conceptualizes the nature of atmospheric turbulence. A brief review of the canonical AO imaging system and H-WFS is followed by a discussion of two image restoration techniques. The problem statement and the proposed solution conclude this chapter.

1.2 Atmospheric turbulence

It is well known that atmospheric turbulence degrades astronomical imaging. The cause, effects, and a measure of atmospheric turbulence are important to understanding the significance of the astronomical imaging problem. The stochastic mechanism causing the turbulence indicates that a statistical solution is needed. A study of the effects tell us that neutralizing the degradation induced by the atmosphere is possible. The measure allows us to analytically characterize the overall turbulence effects with a single parameter.

The fundamental mechanism responsible for the atmospheric turbulence phenomenon is the diurnal heating and cooling of the Earth's surface. Uneven temperature distributions create large scale inhomogeneities (eddies) in the refractive indices of the air [32]. Eddies are homogeneous pockets or regions of air [15]. Kolmogorov asserted that large-scale eddies transmit energy without loss to progressively smaller eddies causing optical turbulence [29]. Hence, the random fluctuations of the refractive indices of air indicate that a stochastic description of the turbulence is needed. As a planar wave front propagates through the atmosphere, the turbulence modulates both the phase and amplitude [8].

The most significant effect of atmospheric turbulence is that it imparts a random tilt onto the wave front [7]. Additionally, the wave front phase perturbations are generally greater than the amplitude distortions. The amplitude distortions appear as scintillation or twinkling. The overall tilt imparted onto the wave front is evident by apparent object motion in the image plane of the telescope. This motion produces the decrease in resolution evident in astronomical imaging. One method of modeling the phase effects on a finer scale is to consider how light rays interact with the atmosphere. When an optical wave front passes through the varying sized eddies, the individual light rays experience varying amounts of retardation. In other words, the planar wave front is *dimpled* and *tilted.*

Fried derived a coherence length measure, r_0, which gauges the overall strength of turbulence-induced perturbations [8]. When the telescope's aperture is greater than r_0, the achievable resolution is equal to that of a telescope of size r_0 in the absence of atmospheric turbulence [8]. For an aperture of dimension r_0, the nearly diffraction limited image appears

to dance around on the detector as the atmosphere evolves [23]. Additionally, r_0 roughly corresponds to the spatial dimensions of a subaperture within which a phase error can be measured and subsequently corrected by a single tilt mirror [29]. The next section discusses the components and operation of adaptive optical imaging systems employed to remove the tilt and dimples from the wave front.

1.3 Adaptive optics

The light emitted or reflected by a celestial object reaching the Earth's atmosphere is essentially a plane wave. As the optical wave front passes through the atmosphere, the planar nature of the wave front is perturbed. The basic premise of adaptive optics is to mechanically deform a reflective surface in the optical train of the telescope to compensate for the atmospheric effects. In other words, the AO system attempts to remove the tilt and dimples in the plane wave front. Specifically, AO imaging systems mitigate the induced phase errors by sensing the perturbations with a WFS and then adding the conjugate phase to the perturbed wave front with a deformable mirror (DM) [17]. The canonical AO system is illustrated in Fig. (1). The controller maps the phase errors sensed by the WFS to actuator commands which modify the DM's figure [25]. The controller employs a wave front reconstruction algorithm to derive the proper commands sent to the actuators [34].

In spite of the amazing achievements of current AO imaging systems [46], these systems are fundamentally limited by many factors. Roggemann and Welsh state that "the problem of obtaining enough light for accurate wave front sensing has been the *most* significant factor limiting the application of AO technology to ground based imaging" [36]. The finite amount of light gathered by the AO system to drive the WFS is strongly linked to performance limiting *shot noise* and *anisoplanatism* [36]. Shot noise or photon noise is due to the random rate of arrival of photons [15, 49]. The technique proposed in this thesis attempts to overcome this limitation by utilizing all of the light collected by the WFS to estimate the wave front slopes over each subaperture. Anisoplanatism results when the angular separation of the object and guide star is greater than the isoplanatic angle, which is on the order of a few arc seconds for

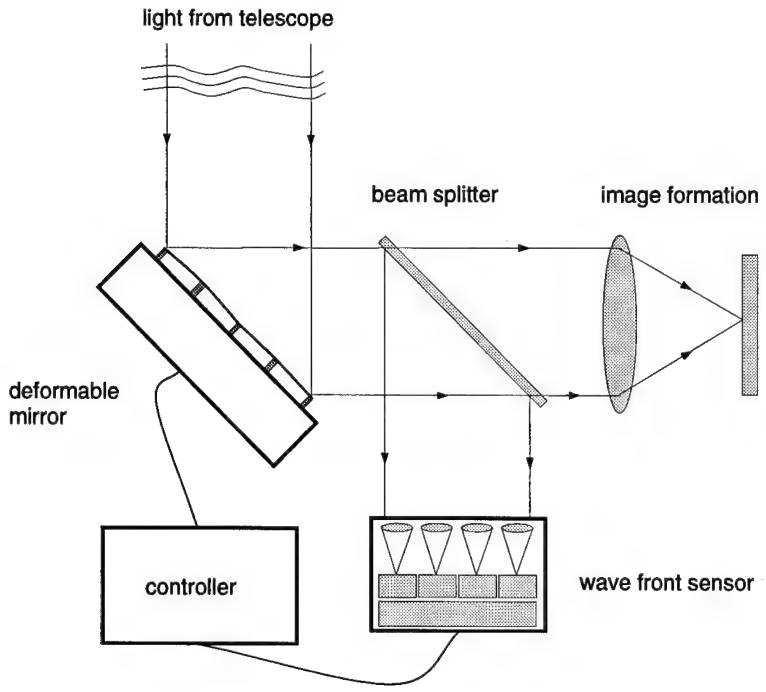

Figure 1. Simplified adaptive optics imaging system with the primary ray paths shown

visible light wavelengths [10]. Other notable limitations include the finite sampling of the wave front by the WFS [48], the limited number of degrees-of-freedom on the DM [33], and delays between sensing and correcting the phase aberrations [20]. The next section discusses a post-processing application which utilizes wave front slope estimation.

1.4 Deconvolution from wave front sensing

The previous section on AO imaging described how astronomical imagery is improved before it is recorded. This section concerns wave front slope estimation in an image post-processing application. Deconvolution from wave front sensing (DWFS) is an image post-processing technique which uses the phase distortion information data collected by the WFS to estimate the point spread function (PSF) [31, 35, 51]. The image is then restored by de-convolving the PSF estimate from the detected image. The deconvolution process effectively increases the resolution in the image. The PSF is a measure of the blurring induced by the optical system, which is dominated by the atmospheric turbulence. Hence, DWFS image

restoration performance is directly linked to the quality of the wave front slope estimation. The technique of applying the DWFS method to images recorded with the aid of an AO system is called compensated DWFS (CDWFS) [47]. In this thesis, a Hartmann-type sensor is employed to gather light from the perturbed wave front.

1.5 Wave front sensing and reconstruction

A Hartmann-type WFS (H-WFS) consists of an array of lenslets with an array of detectors in the back focal plane of each lenslet. Each lenslet forms a subaperture. Each subaperture lenslet forms an intensity distribution or *spot* in the detector. The offset of the spot from the optical axis is termed the shift. The shift is directly proportional to the average wave front tilt or slope within that subaperture [49]. The data recorded by the H-WFS is a composite set of images. This composite image recorded in the back focal plane of the H-WFS is called the WFS image.

Wave front reconstruction involves converting the spot location in each subaperture of the H-WFS to estimate the phase of the wave front. The finite number of photons detected by the H-WFS limits the accuracy of a centroiding procedure. Hence employing the underlying statistical properties of the wave front slope correlations from subaperture to subaperture and the known light levels should increase wave front reconstruction performance.

In summary, wave front reconstruction from measured wave front gradients (slopes) is vital to real-time imaging systems and image post-processing techniques used to restore atmospherically blurred imagery.

1.6 The problem

Develop an algorithm to maximize the use of the data recorded by the H-WFS to optimally estimate the wave front slope over each subaperture of the H-WFS.

1.7 Proposed solution

1.7.1 Introduction. The effects of atmospheric turbulence can be modeled by an equivalent representation of the wave front at the telescope pupil [15], see Fig. (2). The phase aberrations at the pupil are evident in the subimages as spot shifts. Therefore, the data in the short-exposure image collected by the H-WFS is directly related to the aberrated phase of the wave front at the telescope's pupil and can be used to estimate the wave front slope. The composite image recorded in the back focal plane of the H-WFS includes all of the subapertures and is called the *WFS image*. In the following subsections, a conventional method and a maximum likelihood (ML) technique for calculating the spot shifts in the WFS image are discussed and a few comments on the utility of this ML approach are presented.

1.7.2 Conventional method. A conventional method of estimating the wave front slope in the pupil of the imaging system from the WFS data is based on the location of the spot centroid in each subaperture image. The offset of the spot centroid from the expected center position, defined by the optical axis of the lenslet, is due to the random wave front tilt over the corresponding subaperture [24]. A linear phase over the subaperture causes a simple shift in the location of the intensity distribution in the image plane [14]. As atmospheric turbulence is never constant, the lenslet overlays a continuous sequence of images at random offset locations in the image plane. For short-exposure images, the wave front slope over each subaperture is correlated with the neighboring subapertures. This correlation of the wave front tilts is used in an ML estimation approach to estimate the wave front slopes from descrete WFS image data.

1.7.3 Maximum likelihood approach. In order to compute an optimal estimate for the wave front slope, a model of the WFS subaperture slopes in the focal plane of the recording device for each subaperture must be developed. The image created in a specific subaperture is called a *subimage*. Next, the *a priori* knowledge of the wave front slope statistics or correlation properties is factored into the ML technique employed to determine the most likely value for the wave front slope. The following paragraphs briefly sketch the details of this approach.

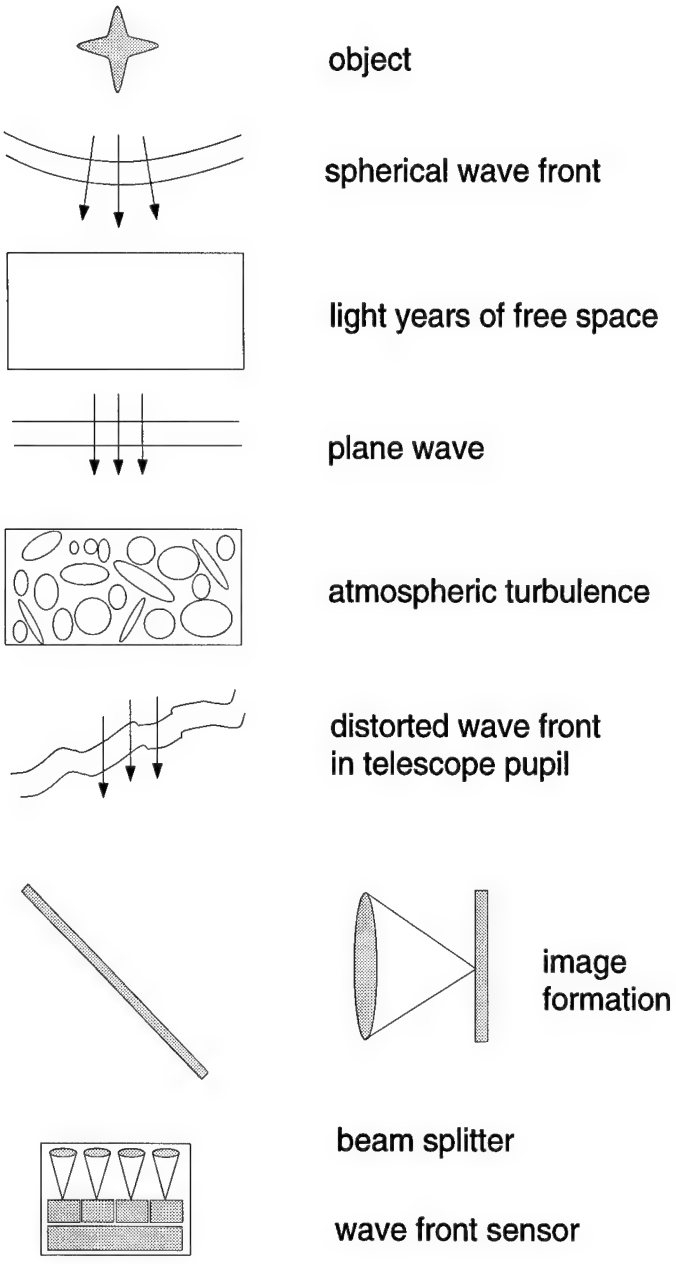

Figure 2. Wave front distortion, wave front sensing, and imaging processes

Using the linear systems framework, the incoherent imaging model equates the image intensity to the convolution of the object intensity and the *point spread function (PSF)* [14]. When a point source is imaged through the turbulent atmosphere, the blurred image is a measure of how the light from the source is spread. The point-source image is the PSF. The detected image is recorded as an array of pixels, thus the detected images are essentially sampled versions of the actual intensity distribution imaged in the back focal plane. Since the detection of light is a random process, it is necessary to employ a stochastic model of the detected image taking into account shot noise. Poisson statistics are well suited to describing shot noise effects [28]. The image can be described by Poisson statistics if we assume that the Poisson parameter or *rate function* is known [30]. The rate function corresponds to the expected intensity of the image. Additionally, the optical detection in one pixel is independent from the detection in another. With these assumptions, the probability density function (PDF) for each subimage is simply a product of PDFs associated with the array of pixels.

The joint PDF of the composite WFS image is derived after developing the detected image model. In the absence of atmospheric turbulence, the PSF and the rate function are known. In reality, the rate function is a random process and optical detection is doubly stochastic [15, 44]. Hence the WFS image model is doubly stochastic. Since phase effects are dominant, the amplitude effects are ignored [7]. The joint PDF corresponding to the WFS image is the product of the conditional PDF calculated for each subimage and a PDF describing the effects of the wave front phase over each subaperture. Modeling the phase perturbation over each subaperture as a *tilt* in the wave front results in a *shift* of the spot in each subimage. The phase tilt is modeled as a zero mean Gaussian random variable [15]. The shift in the spot location is linearly related to the tilt and thus it is also a zero mean Gaussian random variable. Since the phase perturbations across the pupil are known to be spatially and temporally correlated, so are the shifts in each subimage. Because only a single frame of short-exposure imagery is used in the estimation process, the temporal correlation effects are ignored [50]. With knowledge of the H-WFS subaperture geometry, we can deterministically model the rate function and let the randomness be expressed as the shift in the spot centroid

of the subimage. Finally, with the joint PDF determined, the ML approach is used to calculate the wave front slope estimate over each subaperture.

1.7.4 Utility of the maximum likelihood approach. In comparison to conventional techniques which only use the information gathered by the H-WFS, the ML technique derived in this thesis incorporates *a priori* knowledge to improve spot centroid estimation accuracy. In conventional techniques the shift parameter measurement in one subaperture is independent of the calculation in all of the other subapertures. In the presence of excessive shot noise brought on by low light conditions, the true location of the spot may be inaccurately predicted by conventional techniques. When imaging dim objects, the amount of light reaching the optical detectors of the H-WFS may be on par with the background noise. In the ML model-based approach, the joint PDF describing the WFS image employs every pixel in the WFS image to optimally estimate the shift parameter in each subimage using ML estimation. Since the wave front slopes over the subapertures are correlated, the shift parameters are not calculated independent of each other, thus a certain amount of robustness is introduced into the shift parameter calculation.

1.8 Thesis overview

Chapter II covers the theory related to the modeling of the data recorded by the H-WFS. Chapter III covers the slope parameter estimation using the ML technique. Chapter IV presents the properties of the estimators derived in Chapter III. Chapter V gives a summary on the thesis methodology, conclusions, and recommendations.

II. Theory and model derivation

2.1 Introduction

The purpose of this research was to estimate the slope of the distorted wave front with measurements from a Hartmann-type wave front sensor (H-WFS). The wave front slope estimate can then be used to reconstruct the phase distortion modeled at the telescope pupil for image restoration. As the wave front passes through the atmosphere, both its amplitude and phase are perturbed. The most significant distortion is born by the phase [7], implying that scintillation is negligible within a subaperture. Scintillation from subaperture to subaperture is accounted for in the derivation. The slope of the wave front in each subaperture is directly proportional to a shift in the centroid of the detected irradiance distribution [14]. Hence, estimating the centroid shift in each subimage is equivalent to estimating the slope of the wave front over each subaperture.

The fundamentally random nature of optical detection is modeled with Poisson statistics. By allowing the Poisson parameter to be random, the Poisson process can incorporate the stochastic effects induced by the atmosphere [15]. The Poisson parameter is referred to herein as the rate function since it essentially models the rate at which photons arrive at the image detector pixels. The rate function and the expected image are shown to be equivalent quantities. For a point-source object, the rate function is also equivalent to the point spread function (PSF).

The following sections review the principles of maximum likelihood estimation theory for random parameters and proceed to develop the mathematical model describing the incoherent image formation process, including the assumptions made to keep the mathematics tractable. By treating the H-WFS as a collection of imaging systems, the single image model is expanded into the WFS image model. The ultimate goal of this chapter is to derive the joint probability density function, $f_{\mathbf{D},\mathbf{x}_S}(\mathbf{d}, \mathbf{x}_S)$, where $\mathbf{d}$ represents the *data* recorded by the WFS and $\mathbf{x}_S$ represents the subaperture tilt induced spot *shifts*.

2.2 *Maximum likelihood estimation theory*

The most likely value for a random vector parameter $\mathbf{T}$ of a random vector process $\mathbf{Z}$ can be found by employing the maximum likelihood (ML) technique. When the probability density function (PDF) of the random vector parameter $\mathbf{T}$ is known, the ML approach is called maximum *a priori* (MAP) estimation. This review follows the development in Scharf [39].

Consider a random vector $\mathbf{Z} \in \mathcal{R}^N$, parameterized by a random vector $\mathbf{T} \in \mathcal{R}^N$ with joint PDF $f_{\mathbf{Z},\mathbf{T}}(\mathbf{z},\mathbf{t})$. The joint PDF can be written as a product of a conditional PDF and a marginal PDF,

$$f_{\mathbf{Z},\mathbf{T}}(\mathbf{z},\mathbf{t}) = f_{\mathbf{Z}|\mathbf{T}}(\mathbf{z}|\mathbf{t})\, f_{\mathbf{T}}(\mathbf{t}), \tag{1}$$

where the conditional PDF can be used to estimate the parameter $\mathbf{T}$ when the marginal PDF $f_{\mathbf{T}}(\mathbf{t})$ is unknown. For some observed value $\hat{\mathbf{z}}$ of the random variable $\mathbf{Z}$, a particular value for the parameter $\mathbf{t}$ is more probable than some other value. In other words, the PDF is at a maximum value for a particular realization when the ML estimate of the parameter is chosen. This can be stated mathematically as:

$$\hat{\mathbf{t}} = \arg\left[\max_{\mathbf{t}} f_{\mathbf{Z},\mathbf{T}}(\hat{\mathbf{z}},\mathbf{t})\right], \tag{2}$$

where $\hat{\mathbf{t}}$ is known as the *maximum likelihood estimate* of the random variable $\mathbf{T}$. In this sense, the function

$$l(\mathbf{t},\mathbf{z}) = f_{\mathbf{Z},\mathbf{T}}(\hat{\mathbf{z}},\mathbf{t}) \tag{3}$$

evaluated at the observed value $\hat{\mathbf{z}}$ is termed the *likelihood function* and $L(\mathbf{t},\mathbf{z})$ defined as

$$L(\mathbf{t},\mathbf{z}) = \ln\{f_{\mathbf{Z},\mathbf{T}}(\mathbf{z},\mathbf{t})\} \tag{4}$$

is called the *log-likelihood function*. The log-likelihood function is particularly useful for maximizing exponential distributions. The maximum value of a function is determined by differentiating it with respect to the independent variable of interest, setting the result equal

to zero, and then finding the roots of the homogeneous equation. The *score function* is a short-hand notation for the gradient of the log-likelihood function

$$s(\mathbf{t}, \mathbf{z}) = \frac{\partial}{\partial \mathbf{t}} L(\mathbf{t}, \mathbf{z}). \tag{5}$$

Thus the maximum value of the function is determined by solving

$$s(\hat{\mathbf{t}}, \mathbf{z}) = \mathbf{0}, \tag{6}$$

where the ML estimate is denoted $\hat{\mathbf{t}}$. The ML invariance property and several other properties used to characterize the performance of an estimator are discussed in the following subsections.

2.2.1 Invariance property. Using the ML estimate $\hat{\mathbf{t}}$, the ML estimate of a function of the random parameter vector $\mathbf{T}$ can be computed. If the random vector $\mathbf{G} = f(\mathbf{T})$, then by the invariance property of the ML technique, the ML estimate $\hat{\mathbf{g}}$ is $f(\hat{\mathbf{t}})$.

2.2.2 Estimator properties. Two important properties used to gauge the quality or performance of an estimator are the *mean* and *covariance* of estimate $\hat{\mathbf{t}}$ [39]. The estimator mean is used to determine if an estimator is biased. The mean and covariance are defined as

$$\text{mean}(\hat{\mathbf{t}}) = E\{\hat{\mathbf{t}}\} \tag{7}$$

and

$$\text{covariance}(\hat{\mathbf{t}}) = E\left\{[\hat{\mathbf{t}} - E(\hat{\mathbf{t}})][\hat{\mathbf{t}} - E(\hat{\mathbf{t}})]^T\right\}, \tag{8}$$

where the superscript T represents the transpose operator. $\hat{\mathbf{t}}$ is said to be an *unbiased* estimator of the random parameter $\mathbf{T}$ if $E\{\hat{\mathbf{t}}\} - E\{\mathbf{T}\} = \mathbf{0}$. The second property, the covariance of

the estimator, shows up in the definition for the *error covariance matrix*:

$$\begin{aligned} \mathbf{C}_{err} &= E\left\{[\hat{\mathbf{t}} - \mathbf{T}][\hat{\mathbf{t}} - \mathbf{T}]^T\right\} \\ &= E\left\{[\hat{\mathbf{t}} - E(\hat{\mathbf{t}})][\hat{\mathbf{t}} - E(\hat{\mathbf{t}})]^T\right\} + [E(\hat{\mathbf{t}}) - \mathbf{T}][E(\hat{\mathbf{t}}) - \mathbf{T}]^T, \end{aligned} \tag{9}$$

where the first term represents the covariance of $\hat{\mathbf{t}}$ and the second term is the squared bias. The *mean squared error (MSE)* of the estimator $\hat{\mathbf{t}}$ is a sum of the diagonal elements of the error covariance matrix. The $\mathrm{MSE}(\hat{\mathbf{t}})$ is defined as the trace of the error covariance matrix,

$$\mathrm{MSE}(\hat{\mathbf{t}}) = \mathrm{tr}\left(\mathbf{C}_{err}\right) = \sum_{n=1}^{N} E\left\{[\hat{t}_n - T_n,]^2\right\}, \tag{10}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator.

Now that we have discussed the ML estimation technique and a few properties of estimators, we can proceed to the derivation of the joint PDF describing the WFS image. The process begins by deriving the joint PDF corresponding to a single image.

2.3 Single Image

The formation of a single image is a fundamental building block of this thesis. The WFS image is viewed as a collection of single images related by the statistics governing the wave front slopes over the subapertures of the H-WFS. This derivation incorporates the doubly stochastic nature of the optical detection. The following subsections discuss the image formation, the optical detection, the random rate function, and the PDF describing a single image.

2.3.1 Image formation. The propagation of incoherent light through the turbulent atmosphere to the back focal plane of a lens is modeled by [18]

$$i(\vec{x};\phi) = \int_{\vec{x}' \in \mathcal{X}'} o(\vec{x}')s(\vec{x} - \vec{x}';\phi)\,\mathrm{d}\vec{x}', \quad \vec{x} \in \mathcal{X}, \tag{11}$$

where $i(\vec{x};\phi)$ is the resultant image irradiance; $o(\vec{x}')$ is the object of interest; $s(\vec{x}-\vec{x}';\phi)$ is the space-invariant PSF; $\vec{x} \in \mathcal{X}$ is the two-dimensional (2-D) image plane; $\vec{x}' \in \mathcal{X}'$ is the 2-D object plane; and ϕ represents the turbulence induced phase errors. If the angular extent of the object is sufficiently small, on the order of a few arc seconds [32], then the isoplanatic assumption is valid and the PSF can be written as a function of the difference of the coordinates. Hence, incoherent imaging can be modeled as a convolutional process as in Eq. (11). The image is detected using a discrete photon counting device, such as a charged-cooled device (CCD) camera [42]. Since the CCD camera has pixels, the recorded image is actually a sampled version of the continuous image intensity distribution focused onto the image plane Eq. (11). The spatial sampling is implicitly modeled by restricting the 2-D position vector, $\vec{x}$, to the discrete image (sample) space $\mathcal{S}$, by writing Eq. (11) as

$$i[\vec{x};\phi] = \int_{\vec{x}' \in \mathcal{X}'} o(\vec{x}')s(\vec{x}-\vec{x}';\phi)\,\mathrm{d}\vec{x}', \quad \vec{x} \in \mathcal{S}, \tag{12}$$

where the image irradiance is denoted by the symbol $i[\vec{x};\phi]$ and the square brackets indicate the implicit sampling of the continuous process [40]. The terms *image irradiance*, *image intensity distribution*, and *expected image* are synonymously used to refer to the image defined in Eq. (12).

The linear systems approach provides a convenient framework for modeling incoherent imaging. The following development relates the shape of the aperture to the unaberrated PSF. This approach closely follows that of Roggemann and Welsh [36]. In general, the expected image is a function of the object and the PSF. But when the object is a point source, the expected image is equivalent to the PSF as shown by replacing $o(\vec{x})$ in Eq. (12) with $\delta(\vec{x})$, yielding [11]

$$\begin{aligned} i[\vec{x};\phi] &= \int_{\vec{x}' \in \mathcal{X}'} \delta(\vec{x}')s(\vec{x}-\vec{x}';\phi)\,\mathrm{d}\vec{x}', \quad \vec{x} \in \mathcal{S} && (13) \\ &= s[\vec{x};\phi], \quad \vec{x} \in \mathcal{S}. && (14) \end{aligned}$$

Thus the expected image is determined by the atmospheric turbulence and aperture geometry. The following development illustrates how the atmospheric turbulence and aperture geometry influence the resultant image of a point-source object.

For a single aperture, the unaberrated PSF is given by the modulus squared of the coherent impulse response, $h_{opt}(\vec{x})$. Expressed mathematically as

$$s_{opt}(\vec{x}) = |h_{opt}(\vec{x})|^2, \tag{15}$$

where $s_{opt}(\vec{x})$ is the unaberrated PSF. Whereas the PSF maps the object intensity to the image plane, the coherent impulse response maps the complex field of the object to the complex field of the image. The coherent impulse response is equal to the inverse Fourier transform of the coherent transfer function (CTF) and can be expressed as

$$h(\vec{x}) = \mathcal{F}^{-1}[H(\vec{u})] = \int_{\vec{u}\in\mathcal{U}} H(\vec{u}) \exp\{j2\pi(\vec{u}\cdot\vec{x})\}\, \mathrm{d}\vec{u}, \tag{16}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform operator; $H(\vec{u})$ is the CTF; $\vec{u}$ is a 2-D spatial frequency variable; and $\vec{u} \in \mathcal{U}$ represents the 2-D Fourier domain. The shape of the aperture is described by the support of the pupil function, $W_p(\vec{x})$, and is related to the CTF by

$$H(\vec{u}) = W_p(\vec{u}\lambda_{avg} f_l), \tag{17}$$

where λ_{avg} is the average wavelength and f_l is the focal length of the lens. To include the random phase fluctuations of the atmosphere, define the generalized pupil function [14]

$$W(\vec{x};\phi) = \exp\{-j\phi(\vec{x})\}\, W_p(\vec{x}). \tag{18}$$

For this work, the phase $\phi(\vec{x})$ is modeled as

$$\phi(\vec{x}) \approx \vec{a}\cdot\vec{x} \tag{19}$$

where $\cdot$ is the vector dot product. Since the phase slope is most likely not a planar surface, $\vec{a}$ represents the least-mean-squares slope of the wave front. Thus $\phi(\vec{x}) = \vec{a} \cdot \vec{x}$ is a tilted plane within the subaperture. This approximation improves as the ratio r_o/L gets bigger, where r_o is the Fried coherence diameter characterizing the seeing and L is the maximum extent of the subaperture. Rewriting Eq. (18) using the approximation in Eq. (19) yields the following generalized pupil function:

$$W(\vec{x};\vec{a}) = \exp\{-j(\vec{a}\cdot\vec{x})\}\, W_p(\vec{x}). \tag{20}$$

Substituting the generalized pupil function of Eq. (20) into Eq. (17) yields

$$H(\vec{u}) = \exp\{-j(\vec{a}\cdot\vec{u}\lambda_{avg}f_l)\}W_p(\vec{u}\lambda_{avg}f_l). \tag{21}$$

Evaluating Eq. (16) with the CTF of Eq. (21) yields the desired PSF relation [11]

$$\begin{aligned} s(\vec{x};\vec{a}) &= \left|\mathcal{F}^{-1}[\exp\{-j(\vec{a}\cdot\vec{u}\lambda_{avg}f_l)\}\, W_p(\vec{u}\lambda_{avg}f_l)]\right|^2, \\ &= \left|h\left(\vec{x} - \frac{\lambda_{avg}f_l}{2\pi}\vec{a}\right)\right|^2, \\ &= s\left(\vec{x} - \frac{\lambda_{avg}f_l}{2\pi}\vec{a}\right), \end{aligned} \tag{22}$$

where $h\left(\vec{x} - \frac{\lambda_{avg}f_l}{2\pi}\vec{a}\right)$ is the coherent impulse response shifted by $\frac{\lambda_{avg}f_l}{2\pi}\vec{a}$.

2.3.2 Single recorded image. The image recorded by a CCD camera is subject to numerous sources of error. The CCD sensor is subject to photo-conversion noise, readout noise, and various background noises [42, 43]. For astronomical imaging, the photo-conversion and readout noises are the dominant sources of error. *Photo-conversion noise* characterizes the random arrival of photons and is modeled by the Poisson random process $d[\vec{x}]$ [49]. *Readout noise* is also a stochastic process and is often modeled with the Gaussian distribution $n[\vec{x}]$

[30]. The recorded image, $r[\vec{x}]$, is modeled by

$$r[\vec{x}] = d[\vec{x}] + n[\vec{x}], \tag{23}$$

where $d[\vec{x}]$ is the detected image and $n[\vec{x}]$ is a zero mean Gaussian random process [4]. In this research, the detector is assumed to be photon noise limited. In other words, the randomness of the optical detection was assumed to be the dominant source of uncertainty. Hence, the recorded and detected images are the "same."

2.3.3 Single image optical detection model. The randomness inherent in optical detection is accurately modeled by the Poisson random process when the semi-classical model of photo-electric detection is employed. This model is valid because the three defining assumptions of the Poisson point process are satisfied [15]. In short, the probability of observing one photo-event at a time is proportional to the observation time, detection area and intensity; the probability of more than one photo-event per observation time interval is negligible; and the number of photo-events occurring in non-overlapping time intervals are statistically independent. Thus, the discrete image is sufficiently modeled as a collection of independent Poisson random variables [15]. Image detection can be cast as the probability of detecting $d[\vec{x}]$ photo-events at location $\vec{x}$. This detection probability is given by the following conditional Poisson PDF [30]:

$$f_{D[\vec{x}]|\Lambda[\vec{x}]}(d[\vec{x}] \mid \lambda[\vec{x}]) = \frac{\lambda[\vec{x} - \vec{x}_S]^{d[\vec{x}]} \exp\{-\lambda[\vec{x} - \vec{x}_S]\}}{d[\vec{x}]!}, \quad \vec{x} \in \mathcal{S}, \tag{24}$$

where $D[\vec{x}]$ is a random variable characterizing the number of photo events at location $\vec{x} \in \mathcal{S}$; $d[\vec{x}]$ is a realization of $D[\vec{x}]$; $\Lambda[\vec{x}]$ is a random variable characterizing the rate function at location $\vec{x}$; $\lambda[\vec{x} - \vec{x}_S]$ is a specific realization of $\Lambda[\vec{x}]$;

$$\vec{x}_S = \frac{\lambda_{avg} f_l}{2\pi} \vec{a} \tag{25}$$

is the random shift defined in Eq. (22); and $f(\cdot)$ is used to denote the PDF of both continuous and discrete distributions. The rate function represents the average photon count for a particular pixel and can be expressed as

$$\lambda[\vec{x} - \vec{x}_S] = E_{D[\vec{x}]|\Lambda[\vec{x}]}\{D[\vec{x}]\}, \tag{26}$$

where $E\{\cdot\}$ is the expectation operator. The rate function is also equivalent to the expected image irradiance given by Eq. (22). Thus, for a point source, the rate function is equivalent to the PSF, $\lambda[\cdot] = s[\cdot]$. The joint PDF of the photo events for all of the pixels in the image is a product of conditional Poisson PDFs given in Eq. (24),

$$f_{\mathbf{D}|\Lambda}(\mathbf{d}|\lambda) = \prod_{\vec{x}\in\mathcal{S}} \frac{\lambda[\vec{x} - \vec{x}_S]^{d[\vec{x}]} \exp\{-\lambda[\vec{x} - \vec{x}_S]\}}{d[\vec{x}]!}, \tag{27}$$

where $\mathbf{D} = \{D[\vec{x}] : \vec{x} \in \mathcal{S}\}$ and $\Lambda = \{\Lambda[\vec{x}] : \vec{x} \in \mathcal{S}\}$ are vectors of random variables, and $\mathbf{d} = \{d[\vec{x}] : \vec{x} \in \mathcal{S}\}$ and $\lambda = \{\lambda[\vec{x} - \vec{x}_S] : \vec{x} \in \mathcal{S}, \vec{x}_S \in \mathcal{X}\}$ are vector representations of specific realizations. The marginal PDF characterizing the randomness of the rate function is discussed in the following subsection.

2.3.4 The random rate function. In general, the rate function depends on both the imaged object and the shape of the aperture. For a point-source object or a system with a small aperture, the imaged object appears as an unresolved *spot*. The entire spot is assumed to be contained entirely within the image detector area. The relationship between the spot location and compactness is shown in Fig. (3). The intensity distribution of the spot is controlled by the shape of the aperture and the relative position is related to the wave front slope. The spot is offset from the optical-axis (center pixel) by the shift parameter $\vec{x}_S$, which is a realization of the random shift parameter $\vec{X}_S$. Hence the shift incorporates the randomness of the rate function and the conditional PDF of Eq. (27) is re-denoted by $f_{\mathbf{D}|\vec{X}_S}(\mathbf{d}|\vec{x}_S)$. The conditional

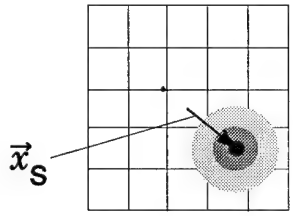

Figure 3. Single image of a compact spot overlayed on the detector pixel array

PDF is repeated for the sake of clarity:

$$f_{\mathbf{D}|\vec{X}_S}(\mathrm{d}|\vec{x}_S) = \prod_{\vec{x}\in\mathcal{S}} \frac{\lambda[\vec{x}-\vec{x}_S]^{d[\vec{x}]}\exp\{-\lambda[\vec{x}-\vec{x}_S]\}}{d[\vec{x}]!}, \tag{28}$$

where $\vec{X}_S$ represents the random bivariate shift variable and $\vec{x}_S$ is a specific realization.

The wave front tilt is manifested as a shift $\vec{x}_S = \frac{\lambda_{avg} f_l}{2\pi}\vec{a}$ of the spot location in the subimage. The PDF of the random shift, $\vec{x}_S$, is modeled as a product of two independent and identically distributed (IID) zero mean Gaussian random variables [8, 15]. The zero mean bivariate normal has the following PDF:

$$f_{\vec{X}_S}(\vec{x}_S) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{|\vec{x}_S|^2}{2\sigma^2}\right\} \tag{29}$$

where σ^2 is the variance of the random shift $\vec{X}_S$ and since $\vec{x}_S = \begin{bmatrix} x_S \\ y_S \end{bmatrix}$, $|\vec{x}_S|^2 = x_S^2 + y_S^2$.

2.3.5 Single image probability density function. By combining the results of the previous subsections, the joint PDF describing a single detected image yields

$$f_{\mathbf{D},\vec{X}_S}(\mathrm{d},\vec{x}_S) = f_{\mathbf{D}|\vec{X}_S}(\mathrm{d}|\vec{x}_S)\, f_{\vec{X}_S}(\vec{x}_S), \tag{30}$$

where $f_{\mathbf{D}|\vec{X}_S}(\mathrm{d}|\vec{x}_S)$ is found using Eq. (28) and $f_{\vec{X}_S}(\vec{x}_S)$ is given in Eq. (29). In the following section, the PDF describing the data recorded by all of the subapertures of the WFS is developed.

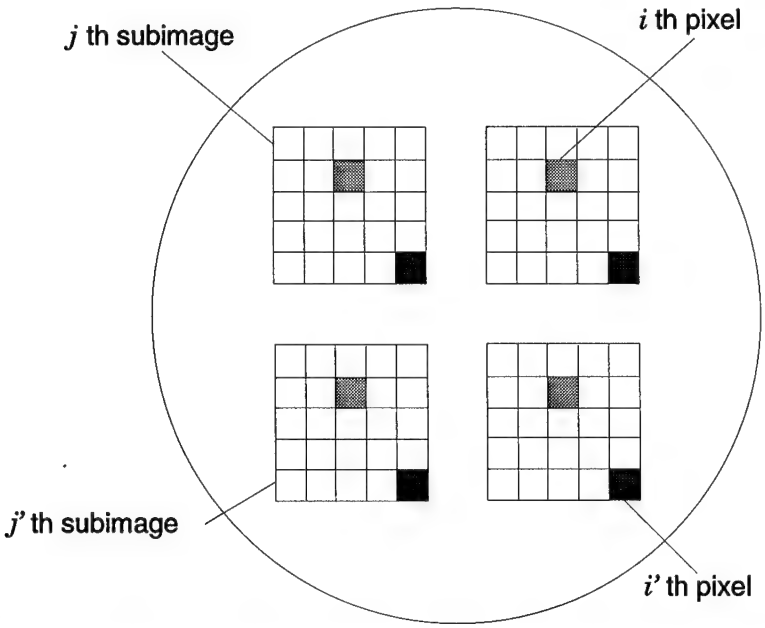

Figure 4. Pixel and subimage labeling system for the wave front sensor image. Note how the pixels labeled by $\vec{x}_i$ and $\vec{x}_{i'}$ are in the same relative positions in all of the subimages.

2.4 Wave front sensor image

The H-WFS can be modeled as a set of correlated diffraction limited imaging systems. The expression for the PDF describing the image in a single aperture is extended to the case of multiple subapertures where the geometry and arrangement of the subapertures is completely arbitrary. The WFS image is defined as a set of J subaperture images or *subimages*, with I pixels in each subimage as shown in Fig. (4). Note that there are $I = 25$ pixels in each of the $J = 4$ subimages in this example.

The pixels in each of the subimages are modeled by a collection of I independent Poisson random variables as in Eq. (24). Let the photon count in a particular pixel be denoted by $d_j[\vec{x}_i]$, where i refers to the pixel location in the jth subimage and $\vec{x}_i \in \mathcal{S}$. The position vector $\vec{x}_i$ locates the ith pixel in any subimage as shown in Fig. (4). The shift parameters $\vec{x}_{S_j}$ and $\vec{x}_{S_{j'}}$ are displayed in Fig. (5).

The joint PDF describing the WFS image is a generalization of Eq. (30). The PDF is a product of the $I \times J$ conditionally Poisson PDFs [30] and the joint PDF of the subaperture

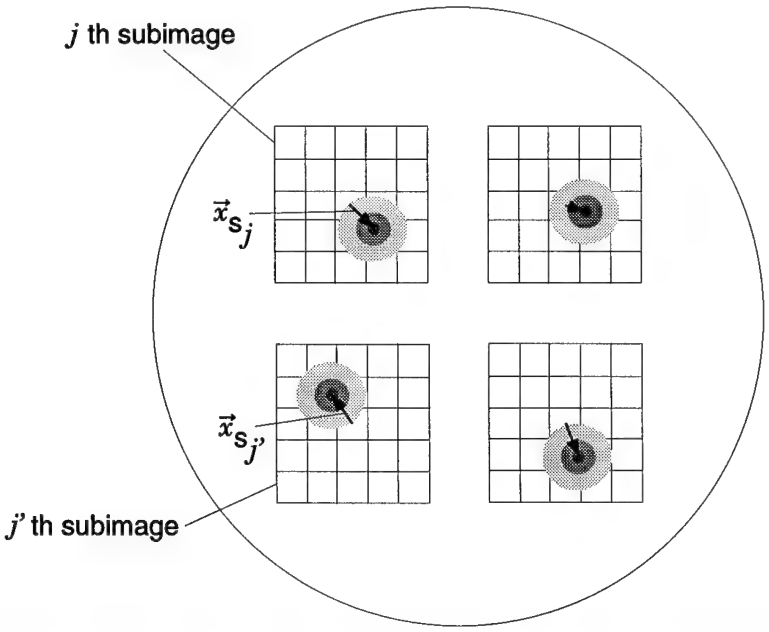

Figure 5. Wave front sensor image with compact spots overlayed on the detector pixel arrays. This figure is intended to show the randomness of the shift parameters.

shifts, $\mathbf{X}_S = (\vec{X}_{S_1}, \ldots, \vec{X}_{S_J})$,

$$f_{\mathbf{D},\mathbf{X}_S}(\mathbf{d}, \mathbf{x}_S) = f_{\mathbf{D}|\mathbf{X}_S}(\mathbf{d}|\mathbf{x}_S)\, f_{\mathbf{X}_S}(\mathbf{x}_S), \tag{31}$$

where

$$f_{\mathbf{D}|\mathbf{X}_S}(\mathbf{d}|\mathbf{x}_S) = \prod_{j=1}^{J} \prod_{i=1}^{I} \frac{\lambda[\vec{x}_i - \vec{x}_{S_j}]^{d_j[\vec{x}_i]} \exp\{-\lambda[\vec{x}_i - \vec{x}_{S_j}]\}}{d_j[\vec{x}_i]!}. \tag{32}$$

The joint PDF of the random shift vector, $\mathbf{X}_S = (\vec{X}_{S_1}, \ldots, \vec{X}_{S_J})$, is derived by extending the bivariate Gaussian of Eq. (29) from one parameter to multiple parameters. The joint PDF of $\mathbf{X}_S$ is the zero mean multivariate normal density

$$f_{\mathbf{X}_S}(\mathbf{x}_S) = (2\pi)^{-J} |\mathbf{R}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{x}_S^T \mathbf{R}^{-1} \mathbf{x}_S\right\}, \tag{33}$$

where $|\mathbf{R}|$ is the determinant of the block correlation matrix $\mathbf{R}$ [28]. The correlation matrix is a block matrix with the following form:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{1,1} & \cdots & \mathbf{R}_{1,J} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{J,1} & \cdots & \mathbf{R}_{J,J} \end{bmatrix}, \tag{34}$$

where each block is given by

$$\mathbf{R}_{jj'} = E\left\{ \begin{bmatrix} x_{S_j} \\ y_{S_j} \end{bmatrix} \begin{bmatrix} x_{S_{j'}} \\ y_{S_{j'}} \end{bmatrix}^T \right\}, \tag{35}$$

and x_{S_j} and y_{S_j} are the coordinates of the spot centroid in the jth subimage [5]. In Eq. (33), $\mathbf{x}_S$ is actually a $2J$-length column vector written as

$$\mathbf{x}_S = \begin{bmatrix} \vec{x}_{S_1} \\ \vdots \\ \vec{x}_{S_J} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} x_{S_1} \\ y_{S_1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} x_{S_J} \\ y_{S_J} \end{bmatrix} \end{bmatrix} \tag{36}$$

which is compatible with the size of the $2J \times 2J$ correlation matrix of Eq. (34). The cross correlation between the x and y coordinate centroid shifts are zero when $j = j'$ since the shifts are assumed to be independent zero mean Gaussian random variables for a single subaperture [28], i.e.

$$\mathbf{R}_{jj} = \begin{bmatrix} \sigma_j^2 & 0 \\ 0 & \sigma_j^2 \end{bmatrix}, \tag{37}$$

where σ_j^2 is the variance of the jth random shift in the both coordinate directions: X_{S_j} and Y_{S_j}. In general, however, the cross correlations $\mathbf{R}_{jj'}$ in Eq. (35) are nonzero. Since $\mathbf{R} \in \mathcal{R}^{2J \times 2J}$ and $\mathbf{R}_{jj'}^T = \mathbf{R}_{j'j}$, $\mathbf{R}^{-1}$ is also real and symmetric [45]. Additionally, $\mathbf{R}$ is positive definite,

which means that $\mathbf{z}^T\mathbf{R}\mathbf{z} > 0, \forall \mathbf{z}$, thus $\mathbf{R}^{-1}$ is also positive definite [28]. Note that the specific correlation properties of the spot shift or equivalently the wave front slope was not essential to this derivation.

2.5 Conclusions

The linear systems framework was used in modeling the incoherent imaging process. Optical detection was modeled by a conditionally Poisson process, where the shift parameter vector is a random process. Equation (31) will be used to find the most likely value for the random vector $\mathbf{X}_S$ for an observed WFS image, $\mathbf{d}$, in the next chapter using ML estimation to optimally determine the wave front slope in the pupil of the telescope.

III. Maximum likelihood parameter estimation

3.1 Introduction

The purpose of the previous chapter was to introduce the estimation theory and model the image captured by the wave front sensor (WFS) with a probability density function (PDF). The PDF describing the WFS image serves the role of the likelihood function in maximum likelihood (ML) estimation. The logarithm is a nondecreasing function of the argument. Taking the logarithm of exponential functions greatly simplifies the maximization procedure. Thus the log-likelihood function will be maximized with respect to the shift parameters. The bulk of the derivation is divided into two parts. In the first part, the score function is derived for an arbitrary rate function and manipulated into a form where the shift parameters for the J subimages are grouped into the shift parameter vector $\mathbf{x}_S$. In the second part, the score function is evaluated for a scaled Gaussian rate function. The Gaussian function adequately models the main lobe of the intensity distribution, which is equivalent to the rate function as stated earlier, provided the volume of the main lobe is approximately equal [52].

3.2 Log-likelihood function

The likelihood function is a scalar defined by Eq. (31) and the log-likelihood function is formed by taking the natural logarithm of both sides of Eq. (31):

$$\begin{aligned} L(\mathbf{x}_S, \mathbf{d}) &= \sum_{j=1}^{J}\sum_{i=1}^{I} d_j[\vec{x}_i]\,\ln\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\} - \lambda[\vec{x}_i - \vec{x}_{S_j}] - \ln\{d_j[\vec{x}_i]!\} \\ &\quad + \ln\left\{(2\pi)^{-J}\right\} + \ln\left\{|\mathbf{R}|^{-1/2}\right\} - \frac{1}{2}\mathbf{x}_S^T\mathbf{R}^{-1}\mathbf{x}_S. \end{aligned} \tag{38}$$

3.3 Deriving the score function

The ML estimate of the shift parameter vector is the root of the score function. The score function is calculated by taking the gradient of the log-likelihood function with respect

to the shift parameters

$$s(\mathbf{x}_S, \mathbf{d}) = \frac{\partial}{\partial \mathbf{x}_S} L(\mathbf{x}_S, \mathbf{d}). \tag{39}$$

Substituting Eq. (38) into Eq. (39) and evaluating the derivative yields

$$\begin{aligned} s(\mathbf{x}_S, \mathbf{d}) &= \frac{\partial}{\partial \mathbf{x}_S} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{I} d_j[\vec{x}_i] \, \ln\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\} - \lambda[\vec{x}_i - \vec{x}_{S_j}] \right\} \\ &\quad - \frac{\partial}{\partial \mathbf{x}_S} \left\{ \frac{1}{2} \mathbf{x}_S^T \mathbf{R}^{-1} \mathbf{x}_S \right\}. \end{aligned} \tag{40}$$

Next, we rewrite the set of terms in the double summation in Eq. (40). The first step involves swapping the order of the summations, so that the inner sum is now over the ith pixel in all J subapertures versus summing the I pixels in each subaperture. Next, the sum over j is expanded and the terms are collected into vectors in Eq. (41). The double sum of Eq. (40) is thus written as

$$\begin{aligned} &\sum_{j=1}^{J} \sum_{i=1}^{I} d_j[\vec{x}_i] \, \ln\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\} - \lambda[\vec{x}_i - \vec{x}_{S_j}] \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} d_j[\vec{x}_i] \, \ln\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\} - \lambda[\vec{x}_i - \vec{x}_{S_j}] \\ &= \sum_{i=1}^{I} d_1[\vec{x}_i] \ln\{\lambda[\vec{x}_i - \vec{x}_{S_1}]\} - \lambda[\vec{x}_i - \vec{x}_{S_1}] + \cdots \\ &\qquad + d_J[\vec{x}_i] \ln\{\lambda[\vec{x}_i - \vec{x}_{S_J}]\} - \lambda[\vec{x}_i - \vec{x}_{S_J}] \\ &= \sum_{i=1}^{I} \ln\left\{\lambda\left[\vec{x}_i \mathbf{1}^T - \mathbf{x}_S^T\right]\right\} \mathbf{d}[\vec{x}_i] - \lambda\left[\vec{x}_i \mathbf{1}^T - \mathbf{x}_S^T\right] \mathbf{1} \end{aligned} \tag{41}$$

where

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}; \quad \vec{x}_i \mathbf{1} = \begin{bmatrix} \vec{x}_i \\ \vdots \\ \vec{x}_i \end{bmatrix}; \tag{42}$$

$\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S]$ and $\ln\{\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S]\}$ are short-hand notations for

$$\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S] = \begin{bmatrix} \lambda[\vec{x}_i - \vec{x}_{S_1}] \\ \vdots \\ \lambda[\vec{x}_i - \vec{x}_{S_J}] \end{bmatrix} \tag{43}$$

and

$$\ln\{\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S]\} = \begin{bmatrix} \ln\{\lambda[\vec{x}_i - \vec{x}_{S_1}]\} \\ \vdots \\ \ln\{\lambda[\vec{x}_i - \vec{x}_{S_J}]\} \end{bmatrix} \tag{44}$$

respectively; $\mathbf{d}[\vec{x}_i]$ is a column vector representing the data in the ith pixel location of each subimage,

$$\mathbf{d}[\vec{x}_i] = \begin{bmatrix} d_1[\vec{x}_i] \\ \vdots \\ d_J[\vec{x}_i] \end{bmatrix}; \tag{45}$$

and $\mathbf{x}_S$ is the shift parameter vector. Also, note that the following notation applies:

$$\lambda\left[\vec{x}_i\mathbf{1}^T - \mathbf{x}_S^T\right] = \left[\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S]\right]^T \tag{46}$$

and

$$\ln\left\{\lambda\left[\vec{x}_i\mathbf{1}^T - \mathbf{x}_S^T\right]\right\} = \left[\ln\{\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S]\}\right]^T. \tag{47}$$

Substituting the result of Eq. (41) into Eq. (40) yields the following simplification for the score function:

$$\begin{aligned} s(\mathbf{x}_S, \mathbf{d}) \;=\; & \frac{\partial}{\partial \mathbf{x}_S}\left\{\sum_{i=1}^{I} \ln\left\{\lambda\left[\vec{x}_i\mathbf{1}^T - \mathbf{x}_S^T\right]\right\} \mathbf{d}[\vec{x}_i] - \lambda\left[\vec{x}_i\mathbf{1}^T - \mathbf{x}_S^T\right]\mathbf{1}\right\} \\ & - \frac{\partial}{\partial \mathbf{x}_S}\left\{\frac{1}{2}\mathbf{x}_S^T\mathbf{R}^{-1}\mathbf{x}_S\right\}, \end{aligned} \tag{48}$$

where the score function is seen to be dependent on the shift parameter vector $\mathbf{x}_S$. To estimate the shift parameters, set the score function equal the zero vector, as in Eq. (5), and then apply the gradient operator to the individual terms in the summation yielding

$$\sum_{i=1}^{I} \frac{\partial}{\partial \mathbf{x}_S} \ln\left\{\lambda\left[\vec{x}_i \mathbf{1}^T - \mathbf{x}_S^T\right]\right\} \mathbf{d}[\vec{x}_i] - \frac{\partial}{\partial \mathbf{x}_S} \lambda\left[\vec{x}_i \mathbf{1}^T - \mathbf{x}_S^T\right] \mathbf{1} - \frac{\partial}{\partial \mathbf{x}_S}\left\{\frac{1}{2}\mathbf{x}_S^T \mathbf{R}^{-1} \mathbf{x}_S\right\} = \mathbf{0}. \tag{49}$$

The gradient of the natural logarithm of $\lambda[\cdot]$ can be simplified in the following manner [16]:

$$\frac{\partial}{\partial \mathbf{x}_S} \ln\{\lambda[\vec{x}_i \mathbf{1} - \mathbf{x}_S]\} = \begin{bmatrix} \frac{\partial}{\partial \vec{x}_{S_1}} \ln\{\lambda[\vec{x}_i - \vec{x}_{S_1}]\} & \cdots & \frac{\partial}{\partial \vec{x}_{S_1}} \ln\{\lambda[\vec{x}_i - \vec{x}_{S_J}]\} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \vec{x}_{S_J}} \ln\{\lambda[\vec{x}_i - \vec{x}_{S_1}]\} & \cdots & \frac{\partial}{\partial \vec{x}_{S_J}} \ln\{\lambda[\vec{x}_i - \vec{x}_{S_J}]\} \end{bmatrix}. \tag{50}$$

Evaluating the partial derivatives of Eq. (50) yields

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{x}_S} \ln\{\lambda[\vec{x}_i \mathbf{1} - \mathbf{x}_S]\} \\ &= \begin{bmatrix} \frac{1}{\lambda[\vec{x}_i - \vec{x}_{S_1}]} \frac{\partial \lambda[\vec{x}_i - \vec{x}_{S_1}]}{\partial \vec{x}_{S_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda[\vec{x}_i - \vec{x}_{S_J}]} \frac{\partial \lambda[\vec{x}_i - \vec{x}_{S_J}]}{\partial \vec{x}_{S_J}} \end{bmatrix} \\ &= \frac{\partial}{\partial \mathbf{x}_S} \lambda[\vec{x}_i \mathbf{1} - \mathbf{x}_S] \, \Lambda_D^{-1}, \end{aligned} \tag{51}$$

where Λ_D^{-1} is a diagonal matrix

$$\Lambda_D^{-1} = \begin{bmatrix} \frac{1}{\lambda[\vec{x}_i - \vec{x}_{S_1}]} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda[\vec{x}_i - \vec{x}_{S_J}]} \end{bmatrix}, \tag{52}$$

and $\frac{\partial}{\partial \mathbf{x}_S}\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S]$ is the gradient of a vector (which is also a diagonal matrix)

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}_S}\lambda[\vec{x}_i\mathbf{1} - \mathbf{x}_S] &= \begin{bmatrix} \frac{\partial}{\partial \vec{x}_{S_1}}\lambda[\vec{x}_i - \vec{x}_{S_1}] & \cdots & \frac{\partial}{\partial \vec{x}_{S_1}}\lambda[\vec{x}_i - \vec{x}_{S_J}] \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \vec{x}_{S_J}}\lambda[\vec{x}_i - \vec{x}_{S_1}] & \cdots & \frac{\partial}{\partial \vec{x}_{S_J}}\lambda[\vec{x}_i - \vec{x}_{S_J}] \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial \vec{x}_{S_1}}\lambda[\vec{x}_i - \vec{x}_{S_1}] & & 0 \\ & \ddots & \\ 0 & & \frac{\partial}{\partial \vec{x}_{S_J}}\lambda[\vec{x}_i - \vec{x}_{S_J}] \end{bmatrix}. \end{aligned} \tag{53}$$

The third term of Eq. (49) depends on the shift vector $\mathbf{x}_S$ and the correlation matrix $\mathbf{R}$. This third term can be reduced using basic linear algebra. The term $\mathbf{x}_S^T\mathbf{R}^{-1}\mathbf{x}_S$ is in pure quadratic form [45]. Given that $\mathbf{x}_S \in \mathcal{R}^{2J}$, $\mathbf{R}^{-1} \in \mathcal{R}^{2J\times 2J}$, and $[\mathbf{R}^{-1}]^T = \mathbf{R}^{-1}$, the gradient of $\mathbf{x}_S^T\mathbf{R}^{-1}\mathbf{x}_S$ with respect to the vector $\mathbf{x}_S$ is [16]

$$\frac{\partial}{\partial \mathbf{x}_S}\mathbf{x}_S^T\mathbf{R}^{-1}\mathbf{x}_S = 2\,\mathbf{R}^{-1}\,\mathbf{x}_S. \tag{54}$$

Combining the results of Eqs. (51) and (54), Eq. (49) can be expressed as

$$\sum_{i=1}^{I} \frac{\partial}{\partial \mathbf{x}_S}\lambda\left[\vec{x}_i\mathbf{1}^T - \mathbf{x}_S^T\right](\Lambda_D^{-1}\mathbf{d}[\vec{x}_i] - \mathbf{1}) - \mathbf{R}^{-1}\mathbf{x}_S = \mathbf{0}. \tag{55}$$

The score function expression in Eq. (55) represents the gradient of the log-likelihood function of the WFS image set equal to zero. Solving Eq. (55) for $\mathbf{x}_S$ yields the ML estimate of the shift parameters. Additional progress cannot be made without assuming a specific form for the rate function. In the following section, a form for the rate function is assumed and the derivation is continued.

3.4 *Solving for the shift parameter vector* $\mathbf{x}_S$

In solving for the shift parameter vector, the generality of the subaperture shape is set aside to make the mathematics tractable. In the following derivation, the intensity profile of the spot in each subimage is assumed to be a Gaussian function. The main goal in modeling the intensity distribution is to accurately estimate the main lobe corresponding to the centroid of the spot. A Gaussian intensity profile scaled by the average number of photo-events detected in the subimage is

$$\lambda[\vec{x}_i - \vec{x}_{S_j}] = \frac{\bar{K}_j l^2}{2\pi\sigma_p^2} \exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma_p^2}\right\}, \tag{56}$$

where $|\vec{x}|^2 = \vec{x}^T\vec{x} = x^2 + y^2$ and the average photon count in the jth subimage is

$$\bar{K}_j = E\left\{\sum_{i=1}^{I} d_j[\vec{x}_i]\right\}, \tag{57}$$

adequately models the intensity distribution in the subaperture. The rate function in Eq. (56) is patterned after Winick's spot intensity distribution [52]. The rate function is modeled by a Gaussian distribution scaled by the average photon count $\bar{K}_j$ and the pixel area l^2, the mean is represented by the shift parameter $\vec{x}_{S_j}$, and the spot size is controlled by the variance σ_p^2. The variance σ_p^2 is dependent on the average imaging wave length λ_{avg}, the focal length of the lenslets f_l [52], and the dimension of the subaperture L. This Gaussian function simplifies the analytical solution for the shift parameters.

Although scintillation effects on the wave front are overshadowed by the phase perturbations [7], the rate function in Eq. (56) is scaled by the photon count in each subimage, $\bar{K}_j$, to allow for scintillation effects from subaperture to subaperture. However, scintillation across each subaperture is assumed to be negligible. In other words, the wave front distortion across each subaperture will be modeled as a random tilting of the plane-wave front.

The analytical solution of Eq. (55) for the rate function defined in Eq. (56) depends on the following assumptions: the spot intensity distribution must be sufficiently compact so that light passing through lenslet j does not bleed over into the detector elements of a neighboring

subaperture j' and large enough so that more than one pixel reports a photo event, hence the following approximation:

$$\sum_{i=1}^{I} (\vec{x}_i - \vec{x}_{S_j}) \exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma^2}\right\} \approx 0 \tag{58}$$

for all j. Figure (5) displays compact spots overlayed onto the detector pixel array. The derivative of the Gaussian rate function defined in Eq. (56) with respect to the jth shift yields a diagonal matrix with the jth element given as

$$\frac{\partial}{\partial \vec{x}_{S_j}} \lambda[\vec{x}_i - \vec{x}_{S_j}] = \frac{\bar{K}_j l^2 (\vec{x}_i - \vec{x}_{S_j})}{2\pi\sigma_p^4} \exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma_p^2}\right\}. \tag{59}$$

Substituting the results of Eq. (59) into Eq. (53) enables the summand of Eq. (55) to be expressed as

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{x}_S} \lambda \left[\vec{x}_i \mathbf{1}^T - \mathbf{x}_S^T\right] \left(\Lambda_D^{-1} \mathbf{d}[\vec{x}_i] - \mathbf{1}\right) \\
&= \begin{bmatrix} \frac{\bar{K}_1 l^2 (\vec{x}_i - \vec{x}_{S_1})}{2\pi\sigma_p^4} \exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_1}|^2}{2\sigma_p^2}\right\} & & 0 \\ & \ddots & \\ 0 & & \frac{\bar{K}_J l^2 (\vec{x}_i - \vec{x}_{S_J})}{2\pi\sigma_p^4} \exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_J}|^2}{2\sigma_p^2}\right\} \end{bmatrix} \\
&\quad \times \left(\begin{bmatrix} \frac{2\pi\sigma_p^2}{\bar{K}_1 l^2} \exp\left\{\frac{|\vec{x}_i - \vec{x}_{S_1}|^2}{2\sigma_p^2}\right\} & & 0 \\ & \ddots & \\ 0 & & \frac{2\pi\sigma_p^2}{\bar{K}_J l^2} \exp\left\{\frac{|\vec{x}_i - \vec{x}_{S_J}|^2}{2\sigma_p^2}\right\} \end{bmatrix} \begin{bmatrix} d_1[\vec{x}_i] \\ \vdots \\ d_J[\vec{x}_i] \end{bmatrix} - \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right).
\end{aligned} \tag{60}$$

Simplifying Eq. (60) gives

$$\frac{\partial}{\partial \mathbf{x}_S}\lambda\left[\vec{x}_i\mathbf{1}^T - \mathbf{x}_S^T\right](\Lambda_D^{-1}\mathbf{d}[\vec{x}_i] - \mathbf{1}) = \begin{bmatrix} \frac{\vec{x}_i - \vec{x}_{S_1}}{\sigma_p^2}d_1[\vec{x}_i] - \frac{\bar{K}_1 l^2(\vec{x}_i - \vec{x}_{S_1})}{2\pi\sigma_p^4}\exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_1}|^2}{2\sigma_p^2}\right\} \\ \vdots \\ \frac{\vec{x}_i - \vec{x}_{S_J}}{\sigma_p^2}d_J[\vec{x}_i] - \frac{\bar{K}_J l^2(\vec{x}_i - \vec{x}_{S_J})}{2\pi\sigma_p^4}\exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_J}|^2}{2\sigma_p^2}\right\} \end{bmatrix}. \tag{61}$$

A sum of vectors is equivalently stated as

$$\sum_{n=1}^{N}\mathbf{z}_n = \begin{bmatrix} \sum_{n=1}^{N} z_{n,1} \\ \vdots \\ \sum_{n=1}^{N} z_{n,M} \end{bmatrix} \quad \text{where } \mathbf{z}_n \in \mathcal{R}^M. \tag{62}$$

The equivalence expressed in Eq. (62) shows how the summation operation in the score function of Eq. (55) can be moved inside the column vector to sum the elements. The score function relation in Eq. (55) can be expressed in the following matrix notation:

$$\begin{bmatrix} \sum_{i=1}^{I}\frac{\vec{x}_i - \vec{x}_{S_1}}{\sigma_p^2}d_1[\vec{x}_i] - \sum_{i=1}^{I}\frac{\bar{K}_1 l^2(\vec{x}_i - \vec{x}_{S_1})}{2\pi\sigma_p^4}\exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_1}|^2}{2\sigma_p^2}\right\} \\ \vdots \\ \sum_{i=1}^{I}\frac{\vec{x}_i - \vec{x}_{S_J}}{\sigma_p^2}d_J[\vec{x}_i] - \sum_{i=1}^{I}\frac{\bar{K}_J l^2(\vec{x}_i - \vec{x}_{S_J})}{2\pi\sigma_p^4}\exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_J}|^2}{2\sigma_p^2}\right\} \end{bmatrix} - \mathbf{R}^{-1}\mathbf{x}_S = \mathbf{0}. \tag{63}$$

Using the approximation stated in Eq. (58), the second summation in each element of Eq. (63) equates to zero. By replacing the summation notation with the inner product of two

vectors, Eq. (63) can be written as

$$\frac{1}{\sigma_p^2}\begin{bmatrix} \mathbf{d}_1^T(\mathbf{x}-\vec{x}_{S_1}\mathbf{1}) \\ \vdots \\ \mathbf{d}_J^T(\mathbf{x}-\vec{x}_{S_J}\mathbf{1}) \end{bmatrix} - \mathbf{R}^{-1}\mathbf{x}_S = \mathbf{0}, \tag{64}$$

where

$$\mathbf{d}_j = \begin{bmatrix} d_j[\vec{x}_1] \\ \vdots \\ d_j[\vec{x}_I] \end{bmatrix} \tag{65}$$

represents the detected photo-events (data) for the jth subimage,

$$\mathbf{x} = \begin{bmatrix} \vec{x}_1 \\ \vdots \\ \vec{x}_I \end{bmatrix} \tag{66}$$

is the pixel position vector, and $\mathbf{1}$ is a column vector of ones. The data is grouped in two manners: $\mathbf{d}_j$ denotes the data vector for all I pixels in the jth subimage and $\mathbf{d}[\vec{x}_i]$ denotes the data vector for the ith pixel in all J subimages.

Simplifying Eq. (64) by separating the data and positioning components from the shift parameters and multiplying through by σ_p^2 yields

$$\begin{bmatrix} \mathbf{d}_1^T\mathbf{x} \\ \vdots \\ \mathbf{d}_J^T\mathbf{x} \end{bmatrix} - \begin{bmatrix} \mathbf{d}_1^T\mathbf{1} & & 0 \\ & \ddots & \\ 0 & & \mathbf{d}_J^T\mathbf{1} \end{bmatrix} \mathbf{x}_S - \sigma_p^2\mathbf{R}^{-1}\mathbf{x}_S = \mathbf{0}. \tag{67}$$

The $\mathbf{d}_j^T\mathbf{1}$ terms on the diagonal is another way of expressing the total number of photons detected in each subimage. Grouping the matrices multiplied by $\mathbf{x}_S$ yields

$$\begin{bmatrix} \mathbf{d}_1^T\mathbf{x} \\ \vdots \\ \mathbf{d}_J^T\mathbf{x} \end{bmatrix} - \left(\begin{bmatrix} \mathbf{d}_1^T\mathbf{1} & & 0 \\ & \ddots & \\ 0 & & \mathbf{d}_J^T\mathbf{1} \end{bmatrix} + \sigma_p^2 \mathbf{R}^{-1} \right) \mathbf{x}_S = \mathbf{0} \tag{68}$$

or in matrix form

$$\mathbf{C}\mathbf{x}_S = \mathbf{m}. \tag{69}$$

The $\mathbf{C}$ (which acts as a **C**orrection factor) and $\mathbf{m}$ (which resembles a **m**oment calculation) are defined as

$$\mathbf{C} = \begin{bmatrix} \mathbf{d}_1^T\mathbf{1} & & 0 \\ & \ddots & \\ 0 & & \mathbf{d}_J^T\mathbf{1} \end{bmatrix} + \sigma_p^2 \mathbf{R}^{-1} = \mathbf{K} + \sigma_p^2 \mathbf{R}^{-1} \tag{70}$$

with the photon count matrix defined as

$$\mathbf{K} = \begin{bmatrix} K_1 & & 0 \\ & \ddots & \\ 0 & & K_J \end{bmatrix} \tag{71}$$

and

$$\mathbf{m} = \begin{bmatrix} \mathbf{d}_1^T\mathbf{x} \\ \vdots \\ \mathbf{d}_J^T\mathbf{x} \end{bmatrix}. \tag{72}$$

Finally, the ML estimate $\hat{\mathbf{x}}_S$ of the shift parameter vector is obtained by solving Eq. (69) for $\mathbf{x}_S$. If the correction factor matrix, $\mathbf{C}$ is nonsingular, then the solution is

$$\hat{\mathbf{x}}_S = \mathbf{C}^{-1}\mathbf{m}. \tag{73}$$

If $\mathbf{C}$ is singular, then calculate $\hat{\mathbf{x}}_S$ using the least mean squares technique [45]:

$$\hat{\mathbf{x}}_S = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{m}. \tag{74}$$

where $\left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T$ is known as a pseudo inverse.

Additionally, by the invariance property of the ML estimate of the shift parameter vector $\mathbf{x}_S$, we can compute the ML estimate of the wave front slope over each subaperture. The slope vector $\mathbf{a}$, is defined as

$$\mathbf{a} = \begin{bmatrix} \vec{a}_1 \\ \vdots \\ \vec{a}_J \end{bmatrix}. \tag{75}$$

The approximate slope for a single aperture, $\vec{a}_j$, is proportional to the shift, $\vec{x}_{S_j}$, or offset of the spot in the subimage by $\vec{a}_j = \dfrac{2\pi}{\lambda_{avg} f_l}\vec{x}_{S_j}$, where λ_{avg} is the average imaging wave length and f_l is the lens focal length. Thus, the ML wave front slope estimate across the telescope pupil is

$$\hat{\mathbf{a}} = \frac{2\pi}{\lambda_{avg} f_l}\hat{\mathbf{x}}_S. \tag{76}$$

3.5 Conclusions

In this chapter, the ML technique was employed to estimate the wave front slope using the PDF constructed in Chapter II to optimally determine the wave front slope in the pupil of the telescope.

IV. Analysis & discussion

4.1 Introduction

Preceding chapters modeled the data recorded by the Hartmann-type wave front sensor (H-WFS) with a probability density function (PDF) and then estimated the shift parameter vector using maximum likelihood (ML) estimation. This chapter explores a few limiting cases for the shift estimator and derives expressions for the first moment and the mean squared error (MSE) of the estimator. The ML shift estimator is shown to be unbiased. The expression for MSE of the estimator is shown to depend on the light level, the correlation properties of the wave front slope statistics, and the classical centroid shift calculation.

With slight loss of generality, the light level in each subaperture is replaced by the average photon count $\bar{K}_j$, defined as

$$\bar{K}_j = E\{K_j\}, \tag{77}$$

where $K_j = \mathbf{d}_j^T \mathbf{1}$. Subsequently, $\mathbf{K}$ is redefine as the average photon count matrix:

$$\mathbf{K} = \begin{bmatrix} \bar{K}_1 & & 0 \\ & \ddots & \\ 0 & & \bar{K}_J \end{bmatrix}. \tag{78}$$

Note that $\mathbf{K}$ is now deterministic. The shift parameter estimate given in Eq. (73) is repeated here for convenience:

$$\hat{\mathbf{x}}_S = \mathbf{C}^{-1}\mathbf{m}, \tag{79}$$

where

$$\mathbf{C} = \mathbf{K} + \sigma_p^2 \mathbf{R}^{-1} \tag{80}$$

is assumed to be nonsingular. Using the invariance property of the ML estimate, the ML estimated slope, $\hat{\mathbf{a}}$, is

$$\hat{\mathbf{a}} = \frac{2\pi}{\lambda_{avg} f_l} \hat{\mathbf{x}}_S, \tag{81}$$

where λ_{avg} is the average imaging wavelength and f_l is the focal length of the lenslets. Since the slope and shifts are related by a constant, the properties of the slope estimator will be explored using the shift parameter estimate, $\hat{\mathbf{x}}_S$. In the next section, the shift vector estimate is shown to be a function of the classical centroid calculation.

4.2 The classical centroid method

The classical centroid method of calculating the spot centroid is adequate for high light. Several researchers have documented the performance of the technique by modeling the optical detection process with the Poisson point process [3, 24, 37]. However, for techniques used in wave front reconstruction, a more accurate technique may be necessary for low light level imaging. Scintillation effects from subaperture to subaperture are accounted for in the model since $\bar{K}_j$ is not necessarily equal to $\bar{K}_{j'}$.

Consider writing Eq. (79) as

$$\hat{\mathbf{x}}_S = \mathbf{C}^{-1}\mathbf{K}\mathbf{K}^{-1}\mathbf{m} = \breve{\mathbf{C}}^{-1}\breve{\mathbf{m}}, \tag{82}$$

where $\breve{\mathbf{m}}$ is equivalent to the classical centroid calculation for the shift vector [24, 37]. The classical centroid offset calculation is

$$\breve{\mathbf{m}} = \begin{bmatrix} \dfrac{\mathbf{d}_1^T\mathbf{x}}{\mathbf{d}_1^T\mathbf{1}} \\ \vdots \\ \dfrac{\mathbf{d}_J^T\mathbf{x}}{\mathbf{d}_J^T\mathbf{1}} \end{bmatrix} = \begin{bmatrix} \dfrac{\vec{m}_1}{\bar{K}_1} \\ \vdots \\ \dfrac{\vec{m}_J}{\bar{K}_J} \end{bmatrix} = \mathbf{K}^{-1}\mathbf{m}, \tag{83}$$

Thus, the ML estimate $\hat{\mathbf{x}}_S$ is a function of the classical centroid shift. From Eq. (82), the inverse of the modified correction factor matrix is defined as

$$\breve{\mathbf{C}}^{-1} = \mathbf{C}^{-1}\mathbf{K}, \tag{84}$$

which implies

$$\breve{\mathbf{C}} = (\mathbf{C}^{-1}\mathbf{K})^{-1} = \mathbf{K}^{-1}\mathbf{C}. \tag{85}$$

Substituting Eq. (70) into Eq. (85) yields the final form of the modified correction factor matrix:

$$\breve{\mathbf{C}} = \mathbf{I} + \sigma_p^2 \mathbf{K}^{-1}\mathbf{R}^{-1}. \tag{86}$$

Harnessing the power of the Sherman-Morrison-Woodbury formula [12], the inverse of $\breve{\mathbf{C}}^{-1}$ becomes

$$\begin{aligned}
\breve{\mathbf{C}}^{-1} &= \mathbf{I}^{-1} - \mathbf{I}^{-1}(\sigma_p^2\mathbf{K}^{-1})\left[\mathbf{I} + \mathbf{R}^{-1}\mathbf{I}^{-1}(\sigma_p^2\mathbf{K}^{-1})\right]^{-1}\mathbf{R}^{-1}\mathbf{I}^{-1}, \\
&= \mathbf{I} - \sigma_p^2\mathbf{K}^{-1}\left(\mathbf{I} + \sigma_p^2\mathbf{R}^{-1}\mathbf{K}^{-1}\right)^{-1}\mathbf{R}^{-1}, \\
&= \mathbf{I} - \left(\mathbf{I} + \frac{1}{\sigma_p^2}\mathbf{R}\mathbf{K}\right)^{-1}.
\end{aligned} \tag{87}$$

Substituting Eq. (87) into Eq. (82) achieves the following relation for the ML shift estimate:

$$\hat{\mathbf{x}}_S = \left[\mathbf{I} - \left(\mathbf{I} + \frac{1}{\sigma_p^2}\mathbf{R}\mathbf{K}\right)^{-1}\right]\breve{\mathbf{m}}. \tag{88}$$

Hence, the ML estimator for the shift vector $\hat{\mathbf{x}}_S$ is equal to the difference of the classical centroid shift calculation $\breve{\mathbf{m}}$ and a correction term based on the relative spot size variance σ_p^2, the light level $\mathbf{K}$ and spot shift correlation properties $\mathbf{R}$. Figure (6) illustrates how the classical centroid shift for the jth subimage, $\breve{m}_j$, represented by the thick dashed line, is modified by

the classical centroid shifts in the other $J-1$ subimages. The $b_{jj'}$ coefficients are elements of

$$\mathbf{B} = \left(\mathbf{I} + \frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}, \tag{89}$$

which is the second term in Eq. (88).

In the following subsections, Eq. (88) is simplified for high or low light levels and when the wave front slopes are uncorrelated.

4.2.1 Bright objects. When imaging bright objects, Eq. (88) reduces to the classical centroiding calculation, implying that the correlation between the slopes over each subaperture are not important. For high light levels, $\mathbf{K}$ is a diagonal matrix of large numbers. Thus

$$\left(\mathbf{I} + \frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1} \approx \mathbf{0}$$

which reduces Eq. (88) to

$$\hat{\mathbf{x}}_S = \left[\mathbf{I} - \left(\mathbf{I} + \frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}\right] \breve{\mathbf{m}} \approx \breve{\mathbf{m}}. \tag{90}$$

When imaging infinitely bright objects, $\hat{\mathbf{x}}_S = \breve{\mathbf{m}}$ which implies that Eq. (90) is exact. Next, the behavior of the estimator is explored for dim objects.

4.2.2 Dim objects. When imaging dim objects, $\mathbf{K}$ is a diagonal matrix of small numbers and thus Eq. (88) reduces to $\hat{\mathbf{x}}_S = \mathbf{0}$. This "no shift" estimate indicates that the photon starved WFS image cannot produce a reliable measurement. To demonstrate this result, let $\mathbf{K}$ be a diagonal matrix of small numbers so that

$$\left(\mathbf{I} + \frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1} \approx \mathbf{I}$$

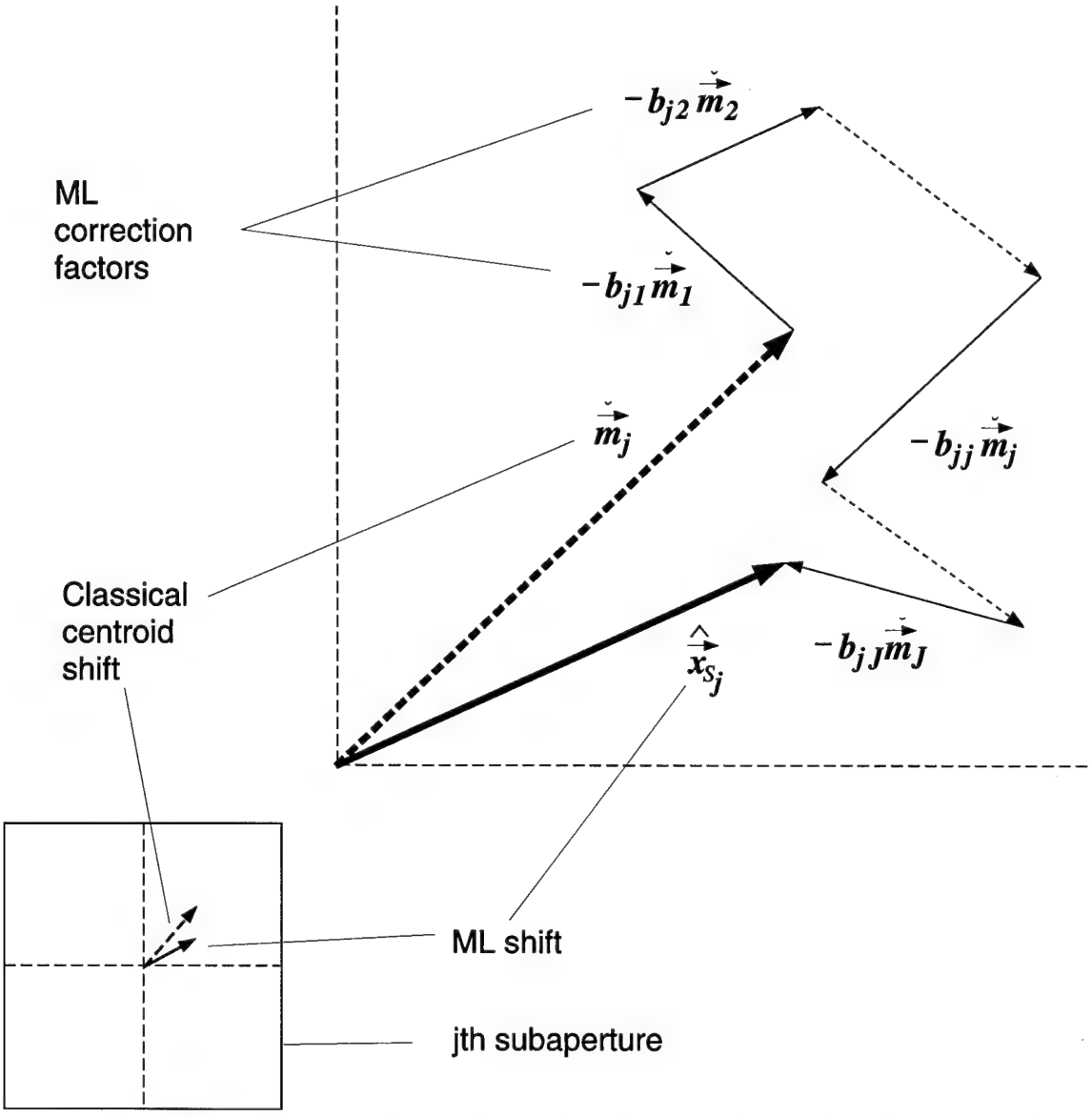

Figure 6. Vector representation of the jth ML shift estimate given in Eq. (88). The tip-to-tail vector addition illustrates how the classical centroid calculation, $\vec{\breve{m}}_j$, is modified by incorporating the centroid shift correlation statistics and light levels to achieve the most likely value for the centroid, $\hat{\vec{x}}_{S_j}$. The vectors $-b_{jj'}\vec{\breve{m}}_j$ for $j' = \{1, \ldots, J\}$ represent the additional *knowledge* provided by the ML theoretic approach. The thin dashed vectors represent the other $J - 4$ unnamed terms modifying the classical centroid shift.

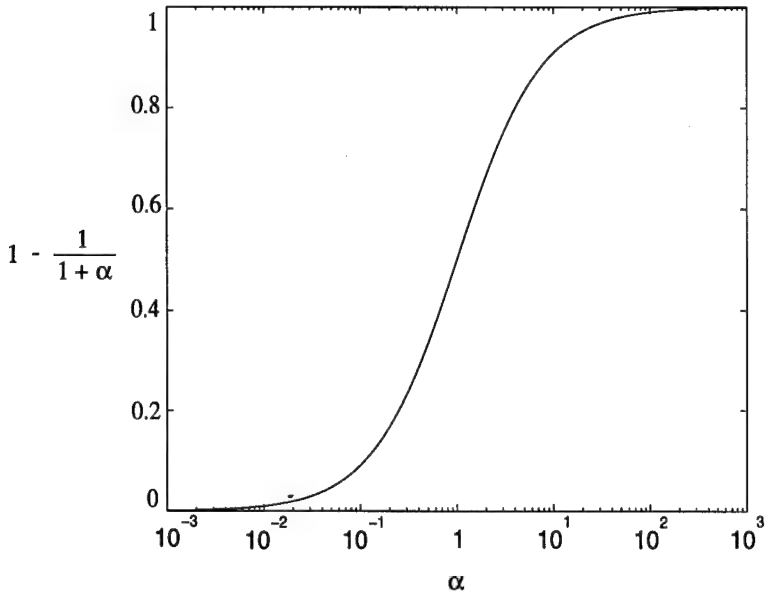

Figure 7. Plot of the factor in Eq. (92)

and thus Eq. (88) becomes

$$\hat{\mathbf{x}}_S = \left[\mathbf{I} - \left(\mathbf{I} + \frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}\right] \breve{\mathbf{m}} \approx (\mathbf{I} - \mathbf{I})\, \breve{\mathbf{m}} = \mathbf{0}. \tag{91}$$

The following case considers the special situation for uncorrelated wave front slopes.

4.2.3 Uncorrelated wave front slopes. When the wave front slopes are uncorrelated, $\mathbf{R}$ becomes a diagonal matrix. Thus all of the matrices in Eq. (88) are diagonal. Hence the jth diagonal block element is given by

$$\hat{\vec{x}}_{S_j} = \left(1 - \frac{1}{1 + \bar{K}_j \sigma_j^2 / \sigma_p^2}\right) \breve{\vec{m}}_j \tag{92}$$

where σ_j^2 is the x or y coordinate shift variance for the jth subimage, $\bar{K}_j$ is the average photon count for the jth subimage, and the factor $1 - \dfrac{1}{1+\alpha}$ is plotted in Fig. (7), with $\alpha = \bar{K}_j \sigma_j^2 / \sigma_p^2$. Note that for high and low light levels, Eq. (92) agrees with the results of the previous subsections.

In the following section, the bias of the shift vector estimate is calculated.

4.3 Is the maximum likelihood shift vector estimate unbiased?

To demonstrate whether or not the ML shift vector estimate, $\hat{\mathbf{x}}_S$, is an unbiased estimator, the first moment of $\hat{\mathbf{x}}_S$ is calculated and is compared to the expected value of the shift vector, $\mathbf{x}_S$. If $E\{\hat{\mathbf{x}}_S\} = E\{\mathbf{x}_S\}$, then $\hat{\mathbf{x}}_S$ is an unbiased estimator of $\mathbf{x}_S$. The first moment of the shift vector estimator is

$$E\{\hat{\mathbf{x}}_S\} = \breve{\mathbf{C}}^{-1}E\{\breve{\mathbf{m}}\} = \breve{\mathbf{C}}^{-1}\mathbf{K}^{-1}E\{\mathbf{m}\}, \tag{93}$$

where $E\{\mathbf{m}\}$ can be expressed as

$$E\{\mathbf{m}\} = E\left\{\begin{bmatrix} \sum_{i=1}^{I}\vec{x}_i d_1[\vec{x}_i] \\ \vdots \\ \sum_{i=1}^{I}\vec{x}_i d_J[\vec{x}_i] \end{bmatrix}\right\} = \begin{bmatrix} \sum_{i=1}^{I}\vec{x}_i E\{d_1[\vec{x}_i]\} \\ \vdots \\ \sum_{i=1}^{I}\vec{x}_i E\{d_J[\vec{x}_i]\} \end{bmatrix}. \tag{94}$$

In Eq. (94) the expected value of $d_j[\vec{x}_i]$ can be written as the following nested conditional expectation relation:

$$E\{d_j[\vec{x}_i]\} = E_{\vec{X}_{S_j}}\left\{E_{D[\vec{x}_i]|\vec{X}_{S_j}}\{d_j[\vec{x}_i]\}\right\}, \tag{95}$$

where the inner expectation can be evaluated using Eq. (26). Hence the inner expectation equates the expected value of the data and the rate function yielding

$$E_{D[\vec{x}_i]|\vec{X}_{S_j}}\{d_j[\vec{x}_i]\} = \lambda[\vec{x}_i - \vec{x}_{S_j}]. \tag{96}$$

Simplifying Eq. (95) with Eq. (96) yields

$$E\{d_j[\vec{x}_i]\} = E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\right\}. \tag{97}$$

and by definition,

$$E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\right\} = \int_{\vec{x}_{S_j}\in\mathcal{X}} \lambda[\vec{x}_i - \vec{x}_{S_j}]\; f_{\vec{X}_{S_j}}(\vec{x}_{S_j})\,\mathrm{d}\vec{x}_{S_j}. \tag{98}$$

The rate function was defined in Eq. (56) as a scaled Gaussian function

$$\lambda[\vec{x}_i - \vec{x}_{S_j}] = \frac{\bar{K}_j l^2}{2\pi\sigma_p^2}\exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma_p^2}\right\}, \tag{99}$$

Equivalent forms for the magnitude squared of a 2-D vector $\vec{x}$ are $|\vec{x}|^2 = x^2 + y^2 = \vec{x}^T\vec{x}$. The PDF is

$$f_{\vec{X}_{S_j}}(\vec{x}_{S_j}) = \frac{1}{2\pi\sigma_j^2}\exp\left\{-\frac{|\vec{x}_{S_j}|^2}{2\sigma_j^2}\right\}, \tag{100}$$

where σ_j^2 is the variance of the jth random shift. Substituting Eqs.(99) and (100) into Eq. (98) yields

$$\begin{aligned}
&E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\right\} \\
&= \int_{\vec{x}_{S_j}\in\mathcal{X}} \frac{\bar{K}_j l^2}{2\pi\sigma_p^2}\exp\left\{-\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma_p^2}\right\}\frac{1}{2\pi\sigma_j^2}\exp\left\{-\frac{|\vec{x}_{S_j}|^2}{2\sigma_j^2}\right\}\mathrm{d}\vec{x}_{S_j} \\
&= \int_{\vec{x}_{S_j}\in\mathcal{X}} \frac{\bar{K}_j l^2}{4\pi^2\sigma_p^2\sigma_j^2}\exp\left\{-\left[\frac{|\vec{x}_i|^2}{2\sigma_p^2} - \frac{2\vec{x}_i^T\vec{x}_{S_j}}{2\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{2\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{2\sigma_j^2}\right]\right\}\mathrm{d}\vec{x}_{S_j}.
\end{aligned} \tag{101}$$

In order to evaluate Eq. (101), the "complete the square" operation is performed on the exponent. For notational convenience, let $\mu_j = \frac{1}{2\sigma_p^2} + \frac{1}{2\sigma_j^2}$, thus the exponent of Eq. (101)

becomes

$$\begin{aligned}
&\frac{|\vec{x}_i|^2}{2\sigma_p^2} - \frac{2\vec{x}_i^T\vec{x}_{S_j}}{2\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{2\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{2\sigma_j^2} \\
&= \mu_j\left(|\vec{x}_{S_j}|^2 - \frac{2}{2\sigma_p^2\mu_j}\vec{x}_i^T\vec{x}_{S_j} + \frac{1}{2\sigma_p^2\mu_j}|\vec{x}_i|^2\right) \\
&= \mu_j\left[\left|\vec{x}_{S_j} - \frac{1}{2\sigma_p^2\mu_j}\vec{x}_i\right|^2 + \left(\frac{1}{2\sigma_p^2\mu_j} - \frac{1}{4\sigma_p^4\mu_j^2}\right)|\vec{x}_i|^2\right], \qquad (102)
\end{aligned}$$

where the square was completed in Eq. (102). Combining Eqs. (101) and (102) yields

$$\begin{aligned}
&E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\right\} \\
&= \frac{\bar{K}_j l^2}{4\pi^2\sigma_p^2\sigma_j^2}\exp\left\{\frac{-1}{2\sigma_p^2}\left(1 - \frac{1}{2\sigma_p^2\mu_j}\right)|\vec{x}_i|^2\right\}\underbrace{\int_{\vec{x}_{S_j}\in\mathcal{X}}\exp\left\{-\mu_j\left|\vec{x}_{S_j} - \frac{1}{2\sigma_p^2\mu_j}\vec{x}_i\right|^2\right\}\mathrm{d}\vec{x}_{S_j}}_{=\pi/\mu_j} \\
&= \frac{\bar{K}_j l^2}{4\pi\mu_j\sigma_p^2\sigma_j^2}\exp\left\{\frac{-1}{2\sigma_p^2}\left(1 - \frac{1}{2\sigma_p^2\mu_j}\right)|\vec{x}_i|^2\right\}, \qquad (103)
\end{aligned}$$

where the constants and $\exp\left\{\frac{-1}{2\sigma_p^2}\left(1 - \frac{1}{2\sigma_p^2\mu_j}\right)|\vec{x}_i|^2\right\}$ have been factored out of the integral and then the integral of $\exp\left\{-\mu_j\left|\vec{x}_{S_j} - \frac{1}{2\sigma_p^2\mu_j}\vec{x}_i\right|^2\right\}$ integrates to $\frac{\pi}{\mu_j}$ because the integrand resembles the product of two Gaussian PDFs [28]. Finally, combining Eqs. (94), (97), and (103) yields the following expression:

$$E\{\mathbf{m}\} = \begin{bmatrix} \frac{\bar{K}_1 l^2}{4\pi\mu_1\sigma_p^2\sigma_1^2}\sum_{i=1}^{I}\vec{x}_i\exp\left\{\frac{-1}{2\sigma_p^2}\left(1 - \frac{1}{2\sigma_p^2\mu_1}\right)|\vec{x}_i|^2\right\} \\ \vdots \\ \frac{\bar{K}_J l^2}{4\pi\mu_J\sigma_p^2\sigma_J^2}\sum_{i=1}^{I}\vec{x}_i\exp\left\{\frac{-1}{2\sigma_p^2}\left(1 - \frac{1}{2\sigma_p^2\mu_J}\right)|\vec{x}_i|^2\right\} \end{bmatrix}. \qquad (104)$$

Each of the J identical elements equates to zero because $\sum_{i=1}^{I}\vec{x}_i\exp\left\{\frac{-1}{2\sigma_p^2}\left(1 - \frac{1}{2\sigma_p^2\mu_j}\right)|\vec{x}_i|^2\right\}$ approximates the expected value of a zero mean bivariate Gaussian random variable. This

approximation was previously stated in Eq. (58) for a centered Gaussian random variable. Hence, $E\{\mathbf{m}\} = \mathbf{0}$ implies $E\{\hat{\mathbf{x}}_S\} = \breve{\mathbf{C}}^{-1}\mathbf{K}^{-1}\mathbf{0}$ and thus

$$E\{\hat{\mathbf{x}}_S\} = E\{\mathbf{x}_S\} = \mathbf{0}. \tag{105}$$

Therefore, $\hat{\mathbf{x}}_S$ is an unbiased estimator since $E\{\hat{\mathbf{x}}_S\}$ and $E\{\mathbf{x}_S\}$ are equal. Another important metric used to quantify the performance of an estimator is the MSE. The MSE for the ML shift vector estimator, $\hat{\mathbf{x}}_S$, is calculated in the next section.

4.4 Mean squared error of the maximum likelihood shift vector estimate

The performance of the estimator is quantified by the MSE metric. The derivation starts with the definition of the MSE.

4.4.1 Start with the definition. By substituting the error covariance matrix of Eq. (9) into the MSE given in Eq. (10) for the ML shift vector, the MSE is written as

$$\begin{aligned} \text{MSE}(\hat{\mathbf{x}}_S) &= \text{tr}\left(E\left\{[\hat{\mathbf{x}}_S - E(\hat{\mathbf{x}}_S)][\hat{\mathbf{x}}_S - E(\hat{\mathbf{x}}_S)]^T\right\}\right) \\ &= \text{tr}\left(E\left\{[\hat{\mathbf{x}}_S][\hat{\mathbf{x}}_S]^T\right\}\right), \end{aligned} \tag{106}$$

since $\hat{\mathbf{x}}_S$ is an unbiased estimator. Substituting the expression for the ML estimator from Eq. (82) into Eq. (106) yields

$$\begin{aligned} \text{MSE}(\hat{\mathbf{x}}_S) &= \text{tr}\left(E\left\{[\breve{\mathbf{C}}^{-1}\breve{\mathbf{m}}][\breve{\mathbf{C}}^{-1}\breve{\mathbf{m}}]^T\right\}\right) \\ &= \text{tr}\left(\breve{\mathbf{C}}^{-1}E\left\{\breve{\mathbf{m}}\breve{\mathbf{m}}^T\right\}\breve{\mathbf{C}}^{-1}\right) \\ &= \text{tr}\left(\breve{\mathbf{C}}^{-1}\mathbf{M}\breve{\mathbf{C}}^{-1}\right), \end{aligned} \tag{107}$$

where $\mathbf{M} = E\left\{\breve{\mathbf{m}}\breve{\mathbf{m}}^T\right\}$ is called the moment covariance matrix.

The MSE relation in Eq. (107) can be written as a sum of terms to show how knowledge of the light level and shift correlation properties influence the MSE of the ML shift estimator. Using Eq. (87) to simplify Eq. (107) yields

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{x}}_S) &= \mathrm{tr}\left(\left[\mathbf{I}-\left(\mathbf{I}+\frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}\right]\mathbf{M}\left[\mathbf{I}-\left(\mathbf{I}+\frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}\right]\right)\\ &= \mathrm{tr}\,(\mathbf{M})-2\,\mathrm{tr}\left(\mathbf{M}\left(\mathbf{I}+\frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}\right)\\ &\quad+\mathrm{tr}\left(\left(\mathbf{I}+\frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}\mathbf{M}\left(\mathbf{I}+\frac{1}{\sigma_p^2}\mathbf{RK}\right)^{-1}\right). \qquad (108)\end{aligned}$$

Thus, the MSE of the shift vector estimate is equal to the sum of the classical centroid MSE calculation, $\mathrm{tr}\,(\mathbf{M})$, and two other terms which incorporate the light level, $\mathbf{K}$, the spot size, σ_p^2, and the shift correlation statistics, $\mathbf{R}$. By incorporating the knowledge of $\mathbf{K}$, σ_p^2, and $\mathbf{R}$, the MSE of the estimator, $\mathrm{MSE}(\hat{\mathbf{x}}_S)$, is reduced. Additionally the MSE of the classical centroid calculation forms an upper bound on the MSE of the ML shift estimate.

Note that for very high levels, the MSE goes to zero! In other words, when there is an abundance of photo-events, the estimator is "exact" in the mean-squared sense. In the following subsections, the moment covariance matrix $\mathbf{M}$ is determined and the result is combined with Eq. (108) to compute the MSE of the shift parameter estimator, $\mathbf{x}_S$.

4.4.2 Calculating the moment covariance matrix. The calculation of the moment covariance in Eq. (108) has been divided into several steps to facilitate the flow of the derivation.

Step 1. Calculation of the moment covariance matrix elements, $\mathbf{M}$, is simplified by using $\breve{\mathbf{m}} = \mathbf{K}^{-1}\mathbf{m}$ and then moving the expectation operator into the matrix,

thus

$$\begin{aligned}
\mathbf{M} &= E\left\{(\mathbf{K}^{-1}\mathbf{m})(\mathbf{K}^{-1}\mathbf{m})^T\right\} \\
&= \mathbf{K}^{-1}E\left\{\mathbf{m}\mathbf{m}^T\right\}\mathbf{K}^{-1} \\
&= \mathbf{K}^{-1}\begin{bmatrix} E\left\{\vec{m}_1\vec{m}_1^T\right\} & \cdots & E\left\{\vec{m}_1\vec{m}_J^T\right\} \\ \vdots & \ddots & \vdots \\ E\left\{\vec{m}_J\vec{m}_1^T\right\} & \cdots & E\left\{\vec{m}_J\vec{m}_J^T\right\} \end{bmatrix}\mathbf{K}^{-1},
\end{aligned} \tag{109}$$

where $E\left\{\vec{m}_j\vec{m}_{j'}^T\right\}$ is a 2×2 matrix.

Step 2. Now, evaluate $E\left\{\vec{m}_j\vec{m}_{j'}^T\right\}$ for all j and j' given

$$\vec{m}_j = \sum_{i=1}^{I} \vec{x}_i d_j[\vec{x}_i]. \tag{110}$$

Computing the elements of $\mathbf{M}$ in Eq. (109) involves evaluating $E\left\{\vec{m}_j\vec{m}_{j'}^T\right\}$, which can be expressed as

$$\begin{aligned}
E\left\{\vec{m}_j\vec{m}_{j'}^T\right\} &= E\left\{\left(\sum_{i=1}^{I}\vec{x}_i d_j[\vec{x}_i]\right)\left(\sum_{i'=1}^{I}\vec{x}_{i'} d_{j'}[\vec{x}_{i'}]\right)^T\right\} \\
&= E\left\{\sum_{i=1}^{I}\sum_{i'=1}^{I}\vec{x}_i\vec{x}_{i'}^T\ d_j[\vec{x}_i]d_{j'}[\vec{x}_{i'}]\right\} \\
&= \sum_{i=1}^{I}\sum_{i'=1}^{I}\vec{x}_i\vec{x}_{i'}^T E\{d_j[\vec{x}_i]d_{j'}[\vec{x}_{i'}]\}.
\end{aligned} \tag{111}$$

Further evaluation of Eq. (111) is performed in two separate calculations, Case I for $j = j'$ and Case II for $j \neq j'$. Case I is related to the diagonal elements of the moment covariance matrix and Case II is related to the off-diagonal elements of the moment covariance matrix.

Case I. When $j = j'$, Eq. (111) becomes

$$E\left\{\vec{m}_j \vec{m}_j^T\right\} = \sum_{i=1}^{I} \sum_{\substack{i'=1 \\ i \neq i'}}^{I} \vec{x}_i \vec{x}_{i'}^T E\{d_j[\vec{x}_i] d_j[\vec{x}_{i'}]\} + \sum_{i=1}^{I} \vec{x}_i \vec{x}_i^T E\left\{(d_j[\vec{x}_i])^2\right\}. \tag{112}$$

Using Eq. (95), $E\{d_j[\vec{x}_i]\}$ can be written as $E_{\vec{X}_{S_j}}\left\{E_{D[\vec{x}_i]|\vec{X}_{S_j}}\{d_j[\vec{x}_i]\}\right\}$, and the expectation over the random shift vector can be factored out of Eq. (112) and written as

$$\begin{aligned} E\left\{\vec{m}_j \vec{m}_j^T\right\} &= E_{\vec{X}_{S_j}}\left\{\sum_{i=1}^{I} \sum_{\substack{i'=1 \\ i \neq i'}}^{I} \vec{x}_i \vec{x}_{i'}^T E_{(D[\vec{x}_i], D[\vec{x}_{i'}])|\vec{X}_{S_j}}\{d_j[\vec{x}_i] d_j[\vec{x}_{i'}]\}\right. \\ &\quad \left. + \sum_{i=1}^{I} \vec{x}_i \vec{x}_i^T E_{D[\vec{x}_i]|\vec{X}_{S_j}}\left\{(d_j[\vec{x}_i])^2\right\}\right\}. \end{aligned} \tag{113}$$

Consider the following facts: $E_{(D[\vec{x}_i], D[\vec{x}_{i'}])|\vec{X}_{S_j}}\{d_j[\vec{x}_i] d_j[\vec{x}_{i'}]\}$ and $E_{D[\vec{x}_i]|\vec{X}_{S_j}}\left\{(d_j[\vec{x}_i])^2\right\}$ are expectations on Poisson random variables, and the variance is equivalent to the first moment of a Poisson random variable [28]. Employing these facts, Eq. (113) can be simplified as

$$\begin{aligned} E\left\{\vec{m}_j \vec{m}_j^T\right\} &= E_{\vec{X}_{S_j}}\left\{\sum_{i=1}^{I} \sum_{i'=1}^{I} \vec{x}_i \vec{x}_{i'}^T E_{D[\vec{x}_i]|\vec{X}_{S_j}}\{d_j[\vec{x}_i]\} E_{D[\vec{x}_{i'}]|\vec{X}_{S_j}}\{d_j[\vec{x}_{i'}]\}\right. \\ &\quad \left. + \sum_{i=1}^{I} \vec{x}_i \vec{x}_i^T E_{D[\vec{x}_i]|\vec{X}_{S_j}}\{d_j[\vec{x}_i]\}\right\}. \end{aligned} \tag{114}$$

Substituting the rate function of Eq. (96) into Eq. (114) and moving the expectation over the shift vector back into the summations yields

$$\begin{aligned} E\left\{\vec{m}_j \vec{m}_j^T\right\} &= \sum_{i=1}^{I} \sum_{i'=1}^{I} \vec{x}_i \vec{x}_{i'}^T E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}] \lambda[\vec{x}_{i'} - \vec{x}_{S_j}]\right\} \\ &\quad + \sum_{i=1}^{I} \vec{x}_i \vec{x}_i^T E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\right\}. \end{aligned} \tag{115}$$

Next, the expected value of the rate functions in the above equation are evaluated. The first moment of the rate function in the single sum was computed in an earlier derivation of

the first moment of the shift estimator. Substituting $\mu_j = \frac{1}{2\sigma_p^2} + \frac{1}{2\sigma_j^2}$ into Eq. (103) yields

$$E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\right\} = \frac{\bar{K}_j l^2}{2\pi(\sigma_p^2+\sigma_j^2)} \exp\left\{-\frac{1}{2(\sigma_p^2+\sigma_j^2)}|\vec{x}_i|^2\right\}. \tag{116}$$

The expected value of $\lambda[\vec{x}_i - \vec{x}_{S_j}]\lambda[\vec{x}_{i'} - \vec{x}_{S_j}]$ in the double sum of Eq. (115) is derived next. Both rate functions are a function of $\vec{X}_{S_j}$, so the definition of the expected value of a function of a single random variable is used to write [28]:

$$E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\lambda[\vec{x}_{i'} - \vec{x}_{S_j}]\right\} = \int_{\vec{x}_{S_j}\in\mathcal{X}} \lambda[\vec{x}_i - \vec{x}_{S_j}]\lambda[\vec{x}_{i'} - \vec{x}_{S_j}]\; f_{\vec{X}_{S_j}}(\vec{x}_{S_j})\,\mathrm{d}\vec{x}_{S_j}. \tag{117}$$

Evaluating Eq. (117) with Eqs. (99) and (100) yields

$$\begin{aligned} &E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\lambda[\vec{x}_{i'} - \vec{x}_{S_j}]\right\} \\ &= \frac{\bar{K}_j^2 l^4}{8\pi^3\sigma_p^4\sigma_j^2}\int_{\vec{x}_{S_j}\in\mathcal{X}} \exp\left\{-\left[\frac{|\vec{x}_i|^2+|\vec{x}_{i'}|^2}{2\sigma_p^2}\right.\right. \\ &\qquad \left.\left. - \frac{2(\vec{x}_i+\vec{x}_{i'})^T\vec{x}_{S_j}}{2\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{2\sigma_j^2}\right]\right\}\mathrm{d}\vec{x}_{S_j}. \end{aligned} \tag{118}$$

For notational convenience, let $\nu_j = \frac{1}{\sigma_p^2} + \frac{1}{2\sigma_j^2}$, and then complete the square on the exponent in Eq. (118) in the following manner:

$$\begin{aligned} &\frac{|\vec{x}_i|^2+|\vec{x}_{i'}|^2}{2\sigma_p^2} - \frac{2(\vec{x}_i+\vec{x}_{i'})^T\vec{x}_{S_j}}{2\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{\sigma_p^2} + \frac{|\vec{x}_{S_j}|^2}{2\sigma_j^2} \\ &= \nu_j\left[|\vec{x}_{S_j}|^2 - \frac{2}{2\sigma_p^2\nu_j}(\vec{x}_i+\vec{x}_{i'})^T\vec{x}_{S_j} + \frac{1}{2\sigma_p^2\nu_j}\left(|\vec{x}_i|^2+|\vec{x}_{i'}|^2\right)\right] \\ &= \nu_j\left[\left|\vec{x}_{S_j} - \frac{1}{2\sigma_p^2\nu_j}(\vec{x}_i+\vec{x}_{i'})\right|^2 - \left(\frac{1}{\sigma_p^2\nu_j}\right)^2|\vec{x}_i+\vec{x}_{i'}|^2 + \frac{1}{2\sigma_p^2\nu_j}\left(|\vec{x}_i|^2+|\vec{x}_{i'}|^2\right)\right]. \end{aligned} \tag{119}$$

Substituting Eq. (119) into Eq. (118) yields

$$
\begin{aligned}
& E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i-\vec{x}_{S_j}]\lambda[\vec{x}_{i'}-\vec{x}_{S_j}]\right\} \\
&= \frac{\bar{K}_j^2 l^4}{8\pi^3\sigma_p^4\sigma_j^2}\exp\left\{\frac{-1}{2\sigma_p^2}\left[|\vec{x}_i|^2+|\vec{x}_{i'}|^2-\frac{2}{\sigma_p^2\nu_j}|\vec{x}_i+\vec{x}_{i'}|^2\right]\right\} \\
&\quad \times \underbrace{\int_{\vec{x}_{S_j}\in\mathcal{X}}\exp\left\{-\nu_j\left|\vec{x}_{S_j}-\frac{1}{2\sigma_p^2\nu_j}(\vec{x}_i+\vec{x}_{i'})\right|^2\right\}\mathrm{d}\vec{x}_{S_j}}_{=\pi/\nu_j} \\
&= \frac{\bar{K}_j^2 l^4}{8\pi^2\nu_j\sigma_p^4\sigma_j^2}\exp\left\{\frac{-1}{2\sigma_p^2}\left[|\vec{x}_i|^2+|\vec{x}_{i'}|^2-\frac{2}{\sigma_p^2\nu_j}|\vec{x}_i+\vec{x}_{i'}|^2\right]\right\}, \qquad (120)
\end{aligned}
$$

where the constants and $\exp\left\{\frac{-1}{2\sigma_p^2}\left[|\vec{x}_i|^2+|\vec{x}_{i'}|^2-\frac{2}{\sigma_p^2\nu_j}|\vec{x}_i+\vec{x}_{i'}|^2\right]\right\}$ have been factored out of the integral and then the integral of $\exp\left\{-\nu_j\left|\vec{x}_{S_j}-\frac{1}{2\sigma_p^2\nu_j}(\vec{x}_i+\vec{x}_{i'})\right|^2\right\}$ integrates to $\frac{\pi}{\nu_j}$ since the integrand resembles a product of two Gaussian PDFs [28]. Finally, substituting $\nu_j=\frac{1}{\sigma_p^2}+\frac{1}{2\sigma_j^2}$ into Eq. (120) yields

$$
\begin{aligned}
& E_{\vec{X}_{S_j}}\left\{\lambda[\vec{x}_i-\vec{x}_{S_j}]\lambda[\vec{x}_{i'}-\vec{x}_{S_j}]\right\} \\
&= \frac{\bar{K}_j^2 l^4}{4\pi^2\sigma_p^2(\sigma_p^2+2\sigma_j^2)}\exp\left\{-\frac{1}{2\sigma_p^2}\left[|\vec{x}_i|^2+|\vec{x}_{i'}|^2-\frac{4\sigma_j^2}{\sigma_p^2+2\sigma_j^2}|\vec{x}_i+\vec{x}_{i'}|^2\right]\right\}. \qquad (121)
\end{aligned}
$$

Equation (121) represents the covariance of two pixels in a single subimage for the doubly stochastic Poisson random variable. Substituting Eqs. (116) and (121) into Eq. (115) yields

$$
\begin{aligned}
& E\left\{\vec{m}_j\vec{m}_j^{\,T}\right\} \\
&= \sum_{i=1}^{I}\sum_{i'=1}^{I}\vec{x}_i\vec{x}_{i'}^{\,T}\frac{\bar{K}_j^2 l^4}{4\pi^2\sigma_p^2(\sigma_p^2+2\sigma_j^2)}\exp\left\{-\frac{1}{2\sigma_p^2}\left[|\vec{x}_i|^2+|\vec{x}_{i'}|^2-\frac{4\sigma_j^2}{\sigma_p^2+2\sigma_j^2}|\vec{x}_i+\vec{x}_{i'}|^2\right]\right\} \\
&\quad + \sum_{i=1}^{I}\vec{x}_i\vec{x}_i^{\,T}\frac{\bar{K}_j l^2}{2\pi(\sigma_p^2+\sigma_j^2)}\exp\left\{-\frac{1}{2(\sigma_p^2+\sigma_j^2)}|\vec{x}_i|^2\right\}. \qquad (122)
\end{aligned}
$$

Now that the covariance elements for $j = j'$ have been evaluated, the off-diagonal elements, $j \neq j'$ which represent the covariance of the pixels for different subimages is undertaken.

Case II. The expected value of $d_j[\vec{x}_i]d_{j'}[\vec{x}_{i'}]$ is the covariance of two pixels in different subimages. Equation (114) is actually a shorthand notation for the following conditional expectation:

$$E\left\{d_j[\vec{x}_i]d_{j'}[\vec{x}_{i'}]\right\} = E_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}\left\{E_{(D[\vec{x}_i],D[\vec{x}_{i'}])|(\vec{X}_{S_j},\vec{X}_{S_{j'}})}\left\{d_j[\vec{x}_i]d_{j'}[\vec{x}_{i'}]\right\}\right\}. \tag{123}$$

Since $d_j[\vec{x}_i]$ and $d_{j'}[\vec{x}_{i'}]$ are conditionally independent Poisson random variables, Eq. (123) can be written as

$$\begin{aligned} E\left\{d_j[\vec{x}_i]d_{j'}[\vec{x}_{i'}]\right\} &= E_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}\left\{E_{D[\vec{x}]|\vec{X}_{S_j}}\left\{d_j[\vec{x}_i]\right\}E_{D[\vec{x}]|\vec{X}_{S_{j'}}}\left\{d_{j'}[\vec{x}_{i'}]\right\}\right\} \\ &= E_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\lambda[\vec{x}_{i'} - \vec{x}_{S_{j'}}]\right\}, \end{aligned} \tag{124}$$

where the first moment of the data is equal to the rate function as shown in Eq. (96). The cross-correlation of the two functions of different random variables in Eq. (124) is given by [28]

$$\begin{aligned} &E_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}\left\{\lambda[\vec{x}_i - \vec{x}_{S_j}]\lambda[\vec{x}_{i'} - \vec{x}_{S_{j'}}]\right\} \\ &= \int_{\vec{x}_{S_j}\in\mathcal{X}}\int_{\vec{x}_{S_{j'}}\in\mathcal{X}}\lambda[\vec{x}_i - \vec{x}_{S_j}]\lambda[\vec{x}_{i'} - \vec{x}_{S_{j'}}]f_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}(\vec{x}_{S_j},\vec{x}_{S_{j'}})\,\mathrm{d}\vec{x}_{S_j}\mathrm{d}\vec{x}_{S_{j'}} \end{aligned} \tag{125}$$

where the joint PDF $f_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}(\vec{x}_{S_j},\vec{x}_{S_{j'}})$ is determined with the aid of Eqs. (33), (34), and (35) for $J = 2$ subapertures. Thus, the joint PDF for the two correlated shift vectors

$$f_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}(\vec{x}_{S_j},\vec{x}_{S_{j'}}) = (2\pi)^{-2}|\tilde{\mathbf{R}}|^{-1/2}\exp\left\{-\frac{1}{2}\tilde{\mathbf{x}}_S^T\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{x}}_S\right\}, \tag{126}$$

where

$$\tilde{\mathbf{x}}_S = \begin{bmatrix} \vec{x}_{S_j} \\ \vec{x}_{S_{j'}} \end{bmatrix}, \tag{127}$$

and $|\tilde{\mathbf{R}}|$ is the determinant of the correlation matrix $\tilde{\mathbf{R}}$ for two subapertures. The correlation matrix is a block matrix with the following form:

$$\tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{R}_{jj} & \mathbf{R}_{jj'} \\ \mathbf{R}_{j'j} & \mathbf{R}_{j'j'} \end{bmatrix} \tag{128}$$

where $\mathbf{R}_{jj'}$ was defined in Eq. (35). Substituting Eqs.(99) and (126) into Eq. (125) yields

$$\begin{aligned} & E_{\vec{x}_{S_j},\vec{x}_{S_{j'}}} \left\{ \lambda[\vec{x}_i - \vec{x}_{S_j}] \lambda[\vec{x}_{i'} - \vec{x}_{S_{j'}}] \right\} \\ &= \int_{\vec{x}_{S_j} \in \mathcal{X}} \int_{\vec{x}_{S_{j'}} \in \mathcal{X}} \bar{K}_j \bar{K}_{j'} \left(\frac{l^2}{2\pi\sigma_p^2} \right)^2 \exp\left\{ -\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma_p^2} \right\} \exp\left\{ -\frac{|\vec{x}_{i'} - \vec{x}_{S_{j'}}|^2}{2\sigma_p^2} \right\} \\ &\qquad \times (2\pi)^{-2} |\tilde{\mathbf{R}}|^{-1/2} \exp\left\{ -\frac{1}{2} \tilde{\mathbf{x}}_S^T \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{x}}_S \right\} \mathrm{d}\vec{x}_{S_j} \mathrm{d}\vec{x}_{S_{j'}}. \end{aligned} \tag{129}$$

Completing the square of the exponents of Eq. (129) is considerably detailed, thus many of the intermediate steps have been omitted from the following derivation. Let $\mathbf{P} = \tilde{\mathbf{R}}^{-1}$ and then expand the exponential terms in Eq. (129) as

$$\begin{aligned} & -\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma_p^2} - \frac{|\vec{x}_{i'} - \vec{x}_{S_{j'}}|^2}{2\sigma_p^2} - \frac{1}{2} \tilde{\mathbf{x}}_S^T \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{x}}_S \\ &= -\frac{1}{2} \left[\frac{\vec{x}_i^T \vec{x}_i}{\sigma_p^2} - \frac{2\vec{x}_i^T \vec{x}_{S_j}}{\sigma_p^2} + \frac{\vec{x}_{S_j}^T \vec{x}_{S_j}}{\sigma_p^2} + \frac{\vec{x}_{i'}^T \vec{x}_{i'}}{\sigma_p^2} - \frac{2\vec{x}_{i'}^T \vec{x}_{S_{j'}}}{\sigma_p^2} + \frac{\vec{x}_{S_{j'}}^T \vec{x}_{S_{j'}}}{\sigma_p^2} \right] \\ &\quad - \frac{1}{2} \left[\vec{x}_{S_j}^T \mathbf{P}_{jj} \vec{x}_{S_j} + \vec{x}_{S_j}^T \mathbf{P}_{jj'} \vec{x}_{S_{j'}} + \vec{x}_{S_{j'}}^T \mathbf{P}_{j'j} \vec{x}_{S_j} + \vec{x}_{S_{j'}}^T \mathbf{P}_{j'j'} \vec{x}_{S_{j'}} \right]. \end{aligned} \tag{130}$$

Note that $\vec{u}^T \mathbf{A} \vec{v} = u_1 a_{1,1} v_1 + u_1 a_{1,2} v_2 + u_2 a_{2,1} v_1 + u_2 a_{2,2} v_2$. The form of the "square" we're completing resembles the following scalar:

$$\begin{aligned} & -\frac{1}{2} \left(\tilde{\mathbf{x}}_S - \tilde{\mathbf{n}} \right)^T \tilde{\mathbf{Q}}^{-1} \left(\tilde{\mathbf{x}}_S - \tilde{\mathbf{n}} \right) \\ = \; & -\frac{1}{2} \left[(\vec{x}_{S_j} - \vec{n}_j)^T \mathbf{Q}_{jj} (\vec{x}_{S_j} - \vec{n}_j)^T + (\vec{x}_{S_{j'}} - \vec{n}_{j'})^T \mathbf{Q}_{j'j} (\vec{x}_{S_j} - \vec{n}_j)^T \right. \\ & \left. + (\vec{x}_{S_j} - \vec{n}_j)^T \mathbf{Q}_{jj'} (\vec{x}_{S_{j'}} - \vec{n}_{j'})^T + (\vec{x}_{S_{j'}} - \vec{n}_{j'})^T \mathbf{Q}_{j'j'} (\vec{x}_{S_{j'}} - \vec{n}_{j'})^T \right], \end{aligned} \tag{131}$$

where $\mathbf{Q} = \tilde{\mathbf{Q}}^{-1}$ is a correlation matrix and $\tilde{\mathbf{n}}$ acts like a mean vector. When Eqs. (130) and (131) are compared, the following relation results:

$$\begin{aligned} & -\frac{|\vec{x}_i - \vec{x}_{S_j}|^2}{2\sigma_p^2} - \frac{|\vec{x}_{i'} - \vec{x}_{S_{j'}}|^2}{2\sigma_p^2} - \frac{1}{2} \tilde{\mathbf{x}}_S^T \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{x}}_S \\ = \; & -\frac{1}{2} \left[\left(\tilde{\mathbf{x}}_S - \tilde{\mathbf{n}} \right)^T \tilde{\mathbf{Q}}^{-1} \left(\tilde{\mathbf{x}}_S - \tilde{\mathbf{n}} \right) - \tilde{\mathbf{n}}^T \tilde{\mathbf{Q}}^{-1} \tilde{\mathbf{n}} + \frac{1}{\sigma_p^2} \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} \right], \end{aligned} \tag{132}$$

where

$$\tilde{\mathbf{Q}}^{-1} = \tilde{\mathbf{R}}^{-1} + \sigma_p^{-2} \mathbf{I}, \tag{133}$$

$$\tilde{\mathbf{n}} = \sigma_p^{-2} \tilde{\mathbf{Q}} \tilde{\mathbf{x}} = \left(\mathbf{I} + \sigma_p^2 \tilde{\mathbf{R}}^{-1} \right)^{-1} \tilde{\mathbf{x}}, \tag{134}$$

and

$$\tilde{\mathbf{x}} = \begin{bmatrix} \vec{x}_i \\ \vec{x}_{i'} \end{bmatrix}. \tag{135}$$

Now the covariance of the pixels is found by substituting the results of Eq. (132) into Eq. (129) to get

$$
\begin{aligned}
& E_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}\left\{\lambda[\vec{x}_i-\vec{x}_{S_j}]\lambda[\vec{x}_{i'}-\vec{x}_{S_{j'}}]\right\} \\
&= \bar{K}_j\bar{K}_{j'}\left(\frac{l^2}{2\pi\sigma_p^2}\right)^2(2\pi)^{-2}|\tilde{\mathbf{R}}|^{-1/2}\exp\left\{-\frac{1}{2\sigma_p^2}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}\right\}\exp\left\{\frac{1}{2}\tilde{\mathbf{n}}^T\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{n}}\right\} \\
&\quad \times \underbrace{\int_{\vec{x}_{S_j}\in\mathcal{X}}\int_{\vec{x}_{S_{j'}}\in\mathcal{X}}\exp\left\{-\frac{1}{2}\left(\tilde{\mathbf{x}}_S-\tilde{\mathbf{n}}\right)^T\tilde{\mathbf{Q}}^{-1}\left(\tilde{\mathbf{x}}_S-\tilde{\mathbf{n}}\right)\right\}\mathrm{d}\vec{x}_{S_j}\mathrm{d}\vec{x}_{S_{j'}}}_{=(2\pi)^2|\tilde{\mathbf{Q}}|^{1/2}} \\
&= \frac{\bar{K}_j\bar{K}_{j'}l^4}{4\pi^2\sigma_p^4}\left(\frac{|\tilde{\mathbf{Q}}|}{|\tilde{\mathbf{R}}|}\right)^{1/2}\exp\left\{-\frac{1}{2\sigma_p^2}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}\right\}\exp\left\{\frac{1}{2}\tilde{\mathbf{n}}^T\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{n}}\right\}. \qquad (136)
\end{aligned}
$$

Equation (136) can be further simplified by substituting the relation for $\tilde{\mathbf{n}}$ found in Eq. (134) and combining the exponentials. Hence, Eq. (136) becomes

$$
\begin{aligned}
& E_{\vec{X}_{S_j},\vec{X}_{S_{j'}}}\left\{\lambda[\vec{x}_i-\vec{x}_{S_j}]\lambda[\vec{x}_{i'}-\vec{x}_{S_{j'}}]\right\} \\
&= \frac{\bar{K}_j\bar{K}_{j'}l^4}{4\pi^2\sigma_p^4}\left(\frac{|\tilde{\mathbf{Q}}|}{|\tilde{\mathbf{R}}|}\right)^{1/2}\exp\left\{-\frac{1}{2\sigma_p^2}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}\right\}\exp\left\{\frac{1}{2\sigma_p^4}\tilde{\mathbf{x}}^T\tilde{\mathbf{Q}}^T\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{Q}}\tilde{\mathbf{x}}\right\}, \\
&= \frac{\bar{K}_j\bar{K}_{j'}l^4}{4\pi^2\sigma_p^4}\left(\frac{|\mathbf{I}+\sigma_p^2\tilde{\mathbf{R}}^{-1}|}{|\sigma_p^2\tilde{\mathbf{R}}^{-1}|}\right)^{1/2}\exp\left\{-\frac{1}{2\sigma_p^2}\tilde{\mathbf{x}}^T\left[\mathbf{I}-\left(\mathbf{I}+\sigma_p^2\tilde{\mathbf{R}}^{-1}\right)^{-1}\right]\tilde{\mathbf{x}}\right\}, \qquad (137)
\end{aligned}
$$

where the equalities $|\tilde{\mathbf{Q}}^{-1}| = |\tilde{\mathbf{Q}}|$ and $\tilde{\mathbf{Q}}^T = \tilde{\mathbf{Q}}$ were used.

Step 3. Now compile the results of Step 2 to compute the block elements of $\mathbf{M}$

$$
[\mathbf{M}]_{jj'} = \frac{1}{\bar{K}_j\bar{K}_{j'}}E\left\{\vec{m}_j\vec{m}_{j'}^T\right\}, \qquad (138)
$$

defined in Eq. (109). Substituting Eq. (122) into Eq. (138) yields

$$[\mathbf{M}]_{jj} = \sum_{i=1}^{I}\sum_{i'=1}^{I} \frac{l^4 \vec{x}_i \vec{x}_{i'}^T}{4\pi^2\sigma_p^2(\sigma_p^2+2\sigma_j^2)} \exp\left\{-\frac{1}{2\sigma_p^2}\left[|\vec{x}_i|^2 + |\vec{x}_{i'}|^2 - \frac{4\sigma_j^2}{\sigma_p^2+2\sigma_j^2}|\vec{x}_i+\vec{x}_{i'}|^2\right]\right\} + \sum_{i=1}^{I} \frac{l^2 \vec{x}_i \vec{x}_i^T}{2\pi(\sigma_p^2+\sigma_j^2)\bar{K}_j} \exp\left\{-\frac{1}{2(\sigma_p^2+\sigma_j^2)}|\vec{x}_i|^2\right\}. \quad (139)$$

The result of Case II is used to calculate the off-diagonal elements of $\mathbf{M}$. Substituting Eq. (137) into Eq. (138) yields

$$[\mathbf{M}]_{jj'} = \sum_{i=1}^{I}\sum_{i'=1}^{I} \frac{l^4 \vec{x}_i \vec{x}_{i'}^T}{4\pi^2\sigma_p^4} \left(\frac{|\mathbf{I}+\sigma_p^2\tilde{\mathbf{R}}^{-1}|}{|\sigma_p^2\tilde{\mathbf{R}}^{-1}|}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma_p^2}\tilde{\mathbf{x}}^T\left[\mathbf{I}-\left(\mathbf{I}+\sigma_p^2\tilde{\mathbf{R}}^{-1}\right)^{-1}\right]\tilde{\mathbf{x}}\right\}. \quad (140)$$

4.4.3 Summary of mean squared error calculation. By definition, the MSE of the ML shift estimate is equal to the trace of the shift estimate covariance matrix. The ML shift estimate was expressed in terms of the classical centroid shift, shift correlation statistics, light levels, and the spot size variance in order to show the relative dependence on these factors. The elements of the moment covariance matrix were then evaluated. To determine the MSE of the ML shift estimate, substitute the results for the moment covariance matrix calculation found in Eqs. (139) and (140) into Eq. (108).

4.5 Conclusions

In this chapter, the ML shift estimator, $\hat{\mathbf{x}}_S = \breve{\mathbf{C}}^{-1}\breve{\mathbf{m}}$, was expressed as a function of the classical centroid calculation, $\breve{\mathbf{m}}$, and *a priori* knowledge of the shift correlation matrix, $\mathbf{R}$, the photon count matrix, $\mathbf{K}$, and the spot size variance, σ_p^2. Several special cases were explored, for instance, at high light levels, the ML shift estimate reduced to the classical centroid calculation. The estimator was found to be unbiased. The MSE of the estimator was then shown to be upper bounded by the MSE of the classical centroid method, which is simply the trace of the moment covariance matrix. Lastly, the elements of the moment covariance matrix were evaluated.

V. Conclusion

5.1 Introduction

Atmospheric turbulence is the primary agent responsible for distorting astronomical imagery and the finite light levels reaching the ground-based imaging system serves to limit the performance of a wave front sensing technique. The blurring of the imagery can be lessened by an adaptive optical (AO) imaging system or image post-processing techniques such as deconvolution from wave front sensing (DWFS). The purpose of this thesis is to derive a maximum likelihood (ML) estimate of the wave front slope.

5.2 Summary of methodology

The methodology employed in this thesis centers around constructing a probability density function (PDF) describing the data collected by a Hartmann-type wave front sensor (WFS). The task begins with the formation of a single image. The single image is then extended into a composite image representing the WFS image. The joint PDF describing WFS is employed by the ML estimation technique to solve for the most likely wave front slopes over each subaperture given the WFS image. The optical detection process is modeled by a conditionally Poisson process, where the Poisson parameter or rate function is itself a random process. The form of the rate function is deterministic, and the stochastic portion of the rate function is represented by the location of the spot in each subimage of the Hartmann-type WFS (H-WFS). Since the slope of the wave front over each subaperture of the H-WFS is linearly related to the spot centroid shift from optical axis of the lenslet, finding the most likely spot shifts is equivalent to estimating the wave front slopes.

5.3 Maximum likelihood shift estimator performance

The ML shift estimator incorporates knowledge of the shift correlation properties, the light level and the spot variance size. This ML estimation-theoretic derivation also contains the classical centroid calculation method as a byproduct. For high and low light levels, the

ML shift estimate reduces to the classical centroid calculation and "no shift" respectively. The estimator was determined to be unbiased. Finally, the MSE of the estimator was shown to be upper bounded by the MSE of the classical centroid method.

5.4 Recommendations

The estimator derived in this thesis has been briefly analyzed. The estimator was shown to be unbiased and the mean squared error was calculated. To further characterize this estimator, simulated WFS images for various H-WFS configurations, wave front slope or spot shift statistics, and light levels must be systematically analyzed. Additionally, the performance of wave front reconstruction techniques using this ML spot position estimator should be explored and then compared and contrasted to minimum variance and least-squares reconstruction techniques such as those analyzed by Roggemann [34].

Bibliography

1. Bell, Trudy E. "Electronics and the stars," *IEEE Spectrum*, 16–24 (August 1995).
2. Bilmont, Marsha F., et al. "Effects of predection atmospheric compensation and post-detection image processing on imagery collected at a ground-based telescope," *SPIE Proceedings: Atmospheric Propagation and Remote Sensing*, *1688*:489–500 (1992).
3. Cao, Genrui and Xin Yu. "Accuracy analysis of a Hartmann-Shack wavefront sensor operated with a faint object," *Optical Engineering*, *25*:586–592 (April 1986).
4. Cederquist, Jack N., et al. "Cramer-Rao lower bound on wavefront sensor error," *Optical Engineering*, *25*:586–592 (April 1986).
5. Chanan, Gary A. "Calculation of wave-front tilt correlations associated with atmospheric turbulence," *Journal of the Optical Society of America A*, *9*:298–301 (February 1992).
6. Davis, Derek K. *Effects of photon noise on unconstrained minimization techniques for iterative blind deconvolution.* MS thesis, AFIT/GSO/ENP/94D-01, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1994.
7. Fried, David L. "Statistics of a geometric representation of wavefront distortion," *Journal of the Optical Society of America*, *55*:1427–1435 (November 1965).
8. Fried, David L. "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *Journal of the Optical Society of America*, *56*:1372–1379 (October 1966).
9. Fried, David L. "Post-detection wavefront distortion compensation," *SPIE Proceedings: Digital Image Recovery and Synthesis*, *828*:127–133 (1987).
10. Gardner, Chester S., et al. "Design and performance analysis of adaptive optical telescopes using laser guide stars," *Proceedings of the IEEE*, *78*:1721–1743 (November 1990).
11. Gaskill, Jack D. *Linear Systems, Fourier Transforms, and Optics*. New York: John Wiley & Sons, 1978.
12. Golub, Gene H. and Charles F. Van Loan. *Matrix Computations* (2nd Edition). Baltimore, Maryland: The Johns Hopkins University Press, 1989.
13. Gonzalez, Rafael C. and Richard E. Woods. *Digital Image Processing*. Reading, Massachusetts: Addison-Wesley Publishing Co., 1993.
14. Goodman, Joseph W. *Introduction to Fourier Optics*. New York: McGraw-Hill, Inc., 1968.
15. Goodman, Joseph W. *Statistical Optics*. New York: John Wiley & Sons, 1985.
16. Graham, Alexander. *Kronecker Products and Matrix Calculus with Applications*. New York: Halsted Press, 1981.

17. Hardy, John W. "Active optics: A new technology for the control of light," *Proceedings of the IEEE, 66*:651–697 (1978).

18. Harris, Sr., James L. "Image evaluation and restoration," *Journal of the Optical Society of America, 56*:569–574 (May 1966).

19. Holmes, Timothy J. "Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach," *Journal of the Optical Society of America A, 9*:1052–1061 (July 1992).

20. Hudgin, Richard H. "Optimal wave-front estimation," *Journal of the Optical Society of America, 67*:378–382 (March 1977).

21. Labeyrie, A. "Attainment of diffraction limited resolution in large telescopes by Fourier analysing speckle patterns in star images," *Astronomy and Astrophysics, 6*:85–87 (1970).

22. Lane, R. G. "Blind deconvolution of speckle images," *Journal of the Optical Society of America A, 9*:1508–1514 (September 1992).

23. Leighton, Robert B. "Concerning the problem of making sharper photographs of the planets," *Scientific American, 194*(6):157–166 (June 1956).

24. McGarty, Terrence P. "The estimation of the "Center of gravity" of a photon density profile in noise," *IEEE Transactions on Aerospace and Electronic Systems, 5*:974–979 (November 1969).

25. Merkle, Fritz. "Adaptive optics," *Physics World*, 33–38 (January 1991).

26. Newton, Isaac. *Opticks*. New York: Dover Publications Inc., 1952.

27. Nisenson, Peter and Richard Barakat. "Partial atmospheric correction with adaptive optics," *Journal of the Optical Society of America A, 4*:2249–2253 (December 1987).

28. Papoulis, Athanasios. *Probability, Random Variables, and Stochastic Processes* (3rd Edition). New York: McGraw-Hill, Inc., 1991.

29. Parenti, Ronald R. "Adaptive optics for astronomy," *Lincoln Laboratory Journal, 5*:93–114 (1992).

30. Paxman, Richard G., et al. "Joint estimation of object and aberrations by using phase diversity," *Journal of the Optical Society of America A, 9*:1072–1085 (July 1992).

31. Primot, J., et al. "Deconvolution from wave-front sensing: a new technique for compensating turbulence-degraded images," *Journal of the Optical Society of America A, 7*:1589–1608 (September 1990).

32. Roddier, François. "The effects of atmospheric turbulence in optical astronomy." *Progress in Optics XIX*, edited by E. Wolf, 283–376, North-Holland, 1981.

33. Roggemann, Michael C. "Limited degree-of-freedom adaptive optics and image reconstruction," *Applied Optics, 30*:4227–4233 (October 1991).

34. Roggemann, Michael C. "Optical performance of fully and partially compensated adaptive optics systems using least-squares and minimum variance phase reconstructors," *Computers & Electrical Engineering*, *18*(6):451–465 (1992).

35. Roggemann, Michael C. and Byron M. Welsh. "Signal-to-noise ratio for astronomical imaging using deconvolution from wavefront sensing," *Applied Optics*, *33*:5400–5414 (August 1994).

36. Roggemann, Michael C. and Byron M. Welsh. *Imaging through the Atmosphere*. Boca Raton, Florida: to be published by CRC Press, 1996.

37. Saleh, Bahaa E. A. "Estimation of the location of an optical object with photodetectors limited by quantum noise," *Applied Optics*, *13*:1824–1827 (August 1974).

38. Scarry, Richard. *Richard Scarry's Simple Simon and Other Rhymes*. Racine, Wisconsin: Western Publishing Company Inc., 1964.

39. Scharf, Louis L. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, Massachusetts: Addison-Wesley Publishing Co., 1991.

40. Schulz, Timothy J. "Estimation-theoretic approach to the deconvolution of atmospherically degraded images with wavefront sensor measurements," *SPIE Proceedings: Digital Image Recovery and Synthesis II*, *2029*:311–320 (1993).

41. Schulz, Timothy J. "Multiframe blind deconvolution of astronomical images," *Journal of the Optical Society of America A*, *10*:1064–1073 (May 1993).

42. Snyder, Donald L., et al. "Image recovery from data acquired with a charge-coupled-device camera," *Journal of the Optical Society of America A*, *10*:1014–1023 (May 1993).

43. Snyder, Donald L., et al. "Compensation for readout noise in CCD images," *Journal of the Optical Society of America A*, *12*:272–283 (February 1995).

44. Snyder, Donald L. and Michael I. Miller. *Random Point Processes in Time and Space* (2nd Edition). New York: Springer-Verlag, 1991.

45. Strang, Gilbert. *Linear Algebra and its Applications* (3rd Edition). Fort Worth, Texas: Harcourt Brace Jovanovich, Inc., 1988.

46. Thompson, Laird A. "Adaptive optics in astronomy," *Physics Today*, 24–31 (December 1994).

47. Thorson, Lori A. *Compensated deconvolution from wavefront sensing*. MS thesis, AFIT/GE/ENG/94D-29, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1994.

48. Wallner, Edward P. "Optimal wave-front correction using slope measurements," *Journal of the Optical Society of America*, *73*:1771–1776 (December 1983).

49. Welsh, Byron M., et al. "Fundamental performance of Hartmann and shearing interferometer wave-front sensor," *Applied Optics*, *34*:4186–4195 (July 1995).

50. Welsh, Byron M. and Chester S. Gardner. "Performance analysis of adaptive-optics systems using laser guide stars and slope sensors," *Journal of the Optical Society of America A*, *6*:1913–1923 (December 1989).

51. Welsh, Byron M. and Michael C. Roggemann. "Signal-to-noise comparison of deconvolution from wave-front sensing with traditional linear and speckle image reconstruction," *Applied Optics*, *34*:2111–2119 (April 1995).

52. Winick, Kim A. "Cramér-Rao lower bounds on the performance of charge-coupled-device optical position estimators," *Journal of the Optical Society of America A*, *3*:1809–1815 (November 1986).

Vita

Captain Scott A. Sallberg was born and raised in the Minneapolis, Minnesota area. After graduating from Spring Lake Park High School in June 1984, he entered the United States Air Force (USAF) the following September. His first assignment was as a main frame computer operator with TRIAD Computer Systems (TRICOMS) at the Strategic Air Command (SAC) Headquarters at Offutt Air Force Base (AFB), Nebraska, from January 1985 to December 1987. While at Offutt AFB, Airman Sallberg began his bachelors of science degree at the University of Nebraska-Omaha (UNO). He traveled to Rolla, Missouri in January 1988 to complete his Bachelors of Science in Electrical Engineering (BSEE) under the sponsorship of the Air Force Institute of Technology (AFIT) Civilian Institution's (CI) Airman Education and Commissioning Program (AECP). Soon after graduating from the University of Missouri-Rolla (UMR) in December 1990, he reported to the Officer Training School (OTS) at Lackland AFB, San Antonio, Texas, in February 1991. He was commissioned a Second Lieutenant in the United States Air Force (USAF) in June 1991. Lieutenant Sallberg's first position was as an Aerodynamic Weapons Design Engineer in the Engineering Division at the National Air Intelligence Center [NAIC, formerly Foreign Technology Division (FTD)] of Wright-Patterson AFB, Ohio. After three years at NAIC, Lieutenant Sallberg was selected to continue his formal education at the Air Force Institute of Technology (AFIT) in May 1994. Captain Sallberg specialized in communications and signal processing while pursuing the Masters of Science in Electrical Engineering (MSEE). His follow-on assignment is with the Avionics directorate of Wright Laboratory, Wright-Patterson AFB, where he will be investigating multi-spectral detection and estimation applications with the Targeting Systems Branch of the Electro Optics Division.

Permanent address: 83 Patty Lou Avenue
Dayton, OH 45459